International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-019-00929-2

ORIGINAL ARTICLE

Selective multi-descriptor fusion for face identification

Xin Wei[1] · **Hui Wang**[1] · **Bryan Scotney**[1] · **Huan Wan**[1]

Received: 16 July 2018 / Accepted: 14 January 2019

Abstract
Over the last 2 decades, face identification has been an active field of research in computer vision. As an important class of image representation methods for face identification, fused descriptor-based methods are known to lack sufficient discriminant information, especially when compared with deep learning-based methods. This paper presents a new face representation method, multi-descriptor fusion (MDF), which represents face images through a combination of multiple descriptors, resulting in hyper-high dimensional *fused descriptor features*. MDF enables excellent performance in face identification, exceeding the state-of-the-art, but it comes with high memory and computational costs. As a solution to the high cost problem, this paper also presents an optimisation method, discriminant ability-based multi-descriptor selection (DAMS), to select a subset of descriptors from the set of 65 initial descriptors whilst maximising the discriminant ability. The MDF face representation, after being refined by DAMS, is named selective multi-descriptor fusion (SMDF). Compared with MDF, SMDF has much smaller feature dimension and is thus usable on an ordinary PC, but still has similar performance. Various experiments are conducted on the CAS-PEAL-R1 and LFW datasets to demonstrate the performance of the proposed methods.

Keywords Face identification · Face recognition · Feature extraction · Feature selection · Objective optimisation

1 Introduction

Over the last 2 decades, face recognition has been an active field of research in computer vision and pattern recognition. Face verification, one form of face recognition that is to verify whether two images are of the same person, has achieved excellent performance in recent years that is better than achieved by humans. According to the latest experimental results on the benchmark LFW dataset [1], the state-of-the-art methods have achieved over 99.50% in accuracy (e.g. DeepID3—99.53% [2], FaceNet—99.63% [3] and Baidu—99.77% [4]), which are above human performance—97.53% [5].

Face identification, another form of face recognition, is to identify the ID of a person, which is however still an unsolved problem. For the same methods, face identification is always less accurate than face verification [6]. In the case of close-set face identification on LFW, the accuracy of DeepID3 [2] and Baidu [4] drop to 98.03% and 96.00%, respectively. And the accuracy of Baidu drops further to 92.09% if it is a open-set identification task on LFW [4]. Therefore, as an important branch of face recognition, face identification is still a challenging task.

Similarly to face verification, face identification has two key processes—face representation and face classification. The state-of-the-art methods for face representation mainly include two types: fused descriptors [7–9] and deep learning-based methods [10–12]. Deep learning-based methods have shown excellent performance in recent years; however, they usually use not only the specified training data but also outside data.[1] By contrast, fused descriptors are also competitive [14, 15], especially when there is no outside data available.

✉ Xin Wei
Wei-X@ulster.ac.uk

Hui Wang
h.wang@ulster.ac.uk

Bryan Scotney
bw.scotney@ulster.ac.uk

Huan Wan
Wan-H@ulster.ac.uk

[1] School of Computing, Ulster University, Belfast, UK

[1] In terms of LFW dataset, "outside data" is defined as the data that is not part of LFW [13]. As the outside data can have a significant impact on experiments, researchers are asked to be specific about whether or what type of outside training data was used to ensure fair comparison of different methods on LFW [13].

Published online: 15 February 2019

In the field of computer vision, an image descriptor is the description of the visual features in an image or video [16]. These visual features can be shape, colour, texture, movement or other abstract features. Among a variety of image descriptors, local binary patterns (LBP) [17–19] is a popular choice and has been studied extensively in the face recognition literature. LBP represents images on the basis of the grey-value differences between neighbouring pixels, which is quite effective in face identification and robust to illumination variance. Some further image descriptors have been proposed in recent years, including chain code-based local descriptor (CCBLD) [20], discriminative embedding method based on the image-to-class distance (I2CDDE) [21], quaternionic local ranking binary pattern (QLRBP) [22] and feature descriptor using entropy rate (FDER) [23]. Different from LBP, which is created based on fixed sampling points in a rectangular or circular region, CCBLD builds a chain by repeatedly searching the maximum or the minimum neighbour around the current position. As a fully supervised local descriptor learning algorithm, I2CDDE tries to learn compact but highly discriminative local feature descriptors based on the image-to-class distances. QLRBP combines Quaternionic Ranking and LBP, and proposes a new quaternionic ranking function to determine the order of two colour pixels. Different from the above-mentioned descriptors, FDER uses a graph structure to describe the image patches generated by the nonsubsampled Contourlet transform, and applies the entropy rate of random walks on the graph to build the final descriptor. In summary, different descriptors manifest in various forms, but all of them are targeted on obtaining highly-discriminative features and invariant features.

In addition, as an assistive technique, feature fusion can alleviate the unreliability brought about by using a single set of features and can introduce additional discriminant information. Feature fusion for image representation can be done at different levels by fusing either the same type of descriptors or different types of descriptors [24]. When image descriptors are combined with a feature fusion technique, they form the so-called fused descriptors. In [25], Nikan et al. proposed a method using feature fusion at the decision level. It divides each facial image into $M * N$ blocks, then uses local phase quantisation (LPQ) and multiscale LBP for feature extraction. In [26], Gao et al. fused local features and global features extracted by gabor wavelet transforms (GWT) and discrete cosine transforms (DCT). These features are fused through a weighted sum at the feature level. Instead of fusing features from different image descriptors, Wei et al. [8] fused features from the same LBP image descriptors obtained with different parameters. The base value in computing an LBP pattern is changed from the intensity of a pixel to the mean intensity of a neighbourhood of the pixel, thus the resulting LBP pattern is less sensitive to noise. Multiple LBP feature vectors are constructed at different spatial scales and combined into a weighted distance function. However, the identification accuracies using such representation methods are usually not the state-of-the-art, due possibly to the lack of sufficient discriminant information in those descriptors. To enhance the discriminative ability of fused descriptors, in this paper dual-cross patterns (DCP) [15] is applied as the basic encoder which aims to maximise the joint Shannon entropy; moreover, we proposed a method called multi-descriptor fusion (MDF) to generate the hyper-high dimensional descriptor features. After that, we try to find a strong and robust classifier for fully utilising the power of MDF.

In face classification, classification methods include C4.5 [27], sparse representation classifier (SRC) [28], k-nearest neighbour classifier (kNN) [29], support vector machine (SVM) [30], neural network (NN) [31], and Naive Bayes [32], among which SRC is a popular choice. SRC works by representing each probe image (i.e. an image that is to be classified) as a linear combination of gallery image samples and optimising the linear combination to minimise the residual. SRC is good for dealing with local facial occlusion (such as random pixel corruption), but it is poor at handling continuous occlusion (due to artefacts such as a hat and sunglasses). Therefore, numerous extensions have been proposed, for example, structured sparse error coding (SSEC) [33], regularized robust coding (RRC) [34], and robust kernel representation with statistical local features (SLF-RKR) [35]. These extensions significantly enhance the robustness of SRC to face occlusion, but they may overfit the occluded training images and decrease the recognition accuracy of SRC on non-occluded data [36]. So Huang et al. [36] proposed kernel extended dictionary (KED), which combines kernel discriminant analysis (KDA) and SRC. It has been shown that KED achieves impressive results on both occluded data and non-occluded data, while using fewer dictionary atoms compared with similar methods like extended sparse representation-based classifier (ESRC) [37] and superposed sparse representation classifier (SSRC) [38]. Due to the excellent performance of KED, if not specified, we take KED as the default classifier for the proposed MDF.

To reduce the memory and computational costs of MDF, we also propose a novel optimisation method called discriminant ability-based multi-descriptor selection (DAMS), which aims to find a specific number of descriptors from the entire descriptor set while maximising the discriminant ability. The new face representation, which is refined by DAMS, is called selective multi-descriptor fusion (SMDF). The main contributions of our work are as follows:

1. We proposed MDF, by which we achieve higher identification accuracy than the state-of-the-art methods. Using dual-cross patterns (DCP) [15] as the basic encoder, MDF fuses a large number of global features and landmark-based local features, and thus it is robust to different types of variance including facial occlusion, illumi-

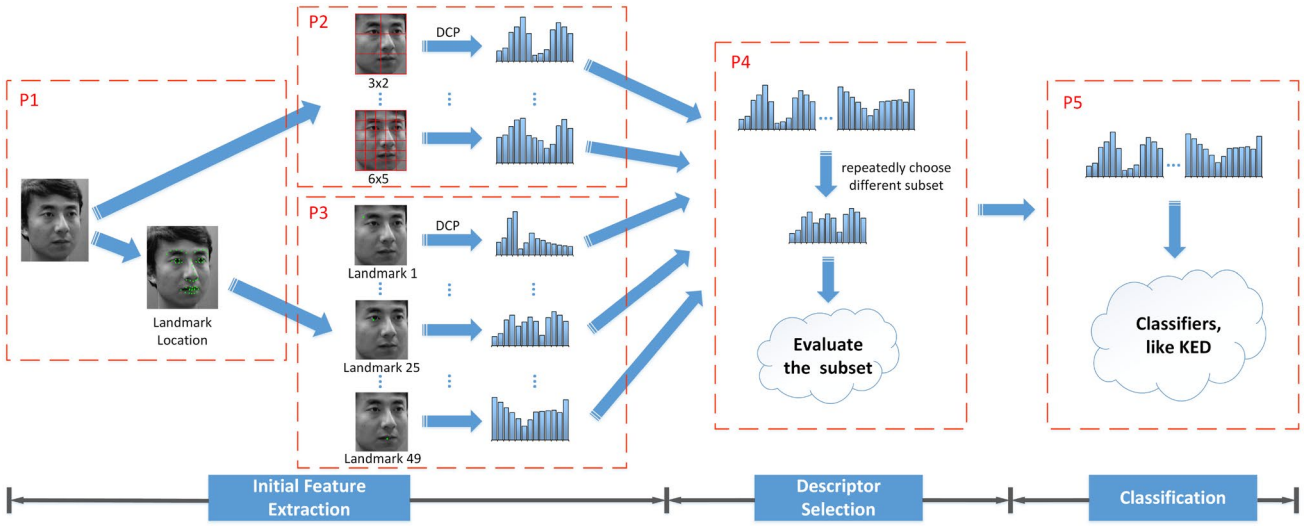


Fig. 1 The pipeline of the proposed methods can be divided into three stages: initial feature extraction, descriptor selection and classification. In the first stage, we employ supervised descent method (SDM) [40] for facial landmark location and extract a large number of global features and landmark-based local features through dual-cross patterns (DCP) [15]. In the second stage, DAMS is applied for descriptor selection. In the final stage, all the features are fused into the classifiers (like KED) for classification. It is worth noting that MDF includes P1, P2, P3 and P5, DAMS refers only to P4, while SMDF consists of P1–P5

nation variance, expression variance, and pose variance. By combining MDF with KED, on the one hand, we take full advantage of the merits of KED, so MDF + KED can cope with the case of one sample per person (OSPP), and just requires a little time to update the whole model when new samples are added into the image gallery. On the other hand, we avoid the demerits, for example, the lack of sufficient discriminant features and robust classifiers. The experimental results on the CAS-PEAL-R1 [39] and LFW dataset [1] show that the performance of MDF is better than DCP, KED and the state-of-the-art methods.

2. We also propose a novel optimisation method called DAMS to reduce the memory and computational costs of MDF. DAMS is designed to be a general optimisation method for searching an optimum subset of feature blocks,[2] where a new objective function is built and a trick based on block matrix operation is utilised to effectively speed up the optimisation process and make it possible in practice. The new face representation, refined by DAMS, is called SMDF. Compared with MDF, SMDF has much smaller feature dimension, which results in a much lower configuration requirement. However, SMDF still achieves excellent performance compared with other methods.

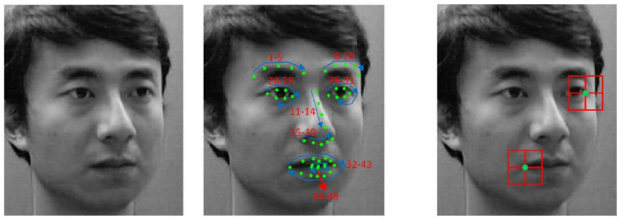
[2] Here, a feature block means a group of features which normally cannot be divided. The features of an instance can consist of many feature blocks. Searching an optimum subset of feature blocks is to find a subset of feature blocks among all feature blocks that can maximise the objective function.

The remainder of the paper is organised as follows. In Sect. 2, firstly we introduce the proposed initial face representation method—MDF; then the optimisation method—DAMS is presented in detail. The experimental results on two commonly used datasets are given in Sect. 3. Finally, we summarize our method and the results in Sect. 4.

2 Selective multi-descriptor fusion

2.1 Initial feature extraction

In this section, the implementation details of MDF are described. As illustrated in Fig. 1, MDF consists of four parts. In part 1, the supervised descent method (SDM) proposed by Xiong et al. [40] is introduced for landmark location. Xiong et al. trained their landmark detector with 66 landmarks on MPIE [41] and LFW-A&C [42] datasets, but only evaluated their landmark detector with 49 of the 66 landmarks on RU-FACS dataset [43], as RU-FACS dataset only provides the ground truths of 49 landmarks. To have a reliable operation, we only use these 49 landmarks in the paper. Alternatively, it may be also possible to use a subset of the 49 landmarks. But at the first stage of our method—initial feature extraction, the control of dimensionality is not our concern. Our top priority at this stage is to generate enough features that contain as much discriminative information as possible. Feature selection will be done at the next stage to reduce the dimensionality. Besides, these 49 landmarks lie on the areas of eyes, nose, mouth, and eyebrow.



(a) (b)

Fig. 2 **a** Landmarks located by SDM. **b** Landmark-based local features

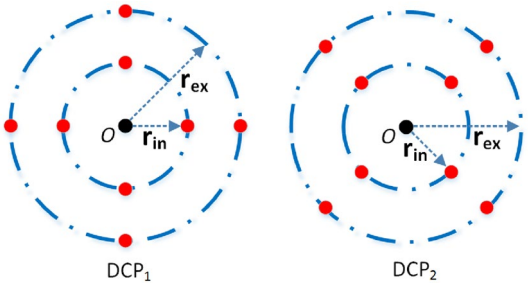


Fig. 3 Dual-cross encoder [15] has two types of patterns—DCP_1 and DCP_2. Each pattern has 8 sampling points distributed on two circles, where r_{in} and r_{ex} are the radii of the inner and exterior circles, respectively

These areas contain very rich features. From our experience, the locations of all these 49 landmarks are important in a facial portion. Neglecting any of these 49 landmarks may lead to a decrease in recognition accuracy. Considering the reasons above, we use all the 49 landmarks in the paper. For each face image, 49 landmarks are located, as shown in Fig. 2a. Being a state-of-the-art method, SDM has very reliable performance on landmark location, which lays a good foundation for subsequent feature extraction.

In part two, DCP [15] is applied to extract the global features for all face images. For each pixel O, Dual-cross encoder is used to obtain the value of each sampling point around it, as illustrated in Fig. 3. Dual-cross encoder [15] includes two types of patterns, namely, DCP_1 and DCP_2. Around each pixel, there are a total of 8 sampling points distributed on two circles. In this paper, the radii of these two circles are denoted by r_{in} and r_{ex}, respectively. Before the global feature extraction, each image is divided into multiple blocks (see P2 of Fig.1), which introduces two parameters—block number of each row (BNR) and block number of each column (BNC). In order to obtain sufficient global DCP features, we adjust the four variables BNR, BNC, r_{in} and r_{ex}. Then the corresponding DCP histogram is calculated for each block; the DCP histograms under the same variable combination are concatenated to form one large feature histogram. In this way, we generate 16 large feature histograms[3] as the global features for each face image. Here, choosing 16 global descriptors is mainly due to two aspects. Firstly, the combination of the global descriptors and the local descriptors requires that their total dimensionalities should be of the same order of magnitude. Ample evidence shows that only in this way can the combination be effective [15, 44–46]. Secondly, the server we use to run the experiments has a memory of 24GB. And the maximum memory usage of MDF is 22GB after applying 16 global descriptors. We choose to maximise the usage of memory and generate as many global descriptors as possible so as to obtain more discriminative features. For the two reasons above, we finally choose to use 16 global descriptors.

In part three, the extraction process of the landmark-based local features is described. As illustrated in Fig. 2a, 49 landmarks are located for each face image. Centred on each landmark, we define a square patch which is divided into $N * N$ non-overlapping blocks as shown in Fig. 2b, where $N = 2$ in Fig. 2b. One DCP histogram is calculated for each block. Hence, $N * N$ DCP histograms are calculated for each landmark. All these $N * N$ DCP histograms are concatenated to build a large histogram as the local DCP features corresponding to this landmark point. From landmark 1 to landmark 49, a total of 49 large DCP histograms are extracted for one face image. Benefiting from the maturity of facial landmark location techniques in recent years, the local features extracted in this part are robust to pose variance, expression variance and distance variance.

In part four, which is P5 in Fig. 1 (here, we skip P4 in Fig. 1, as P4 is an optional module and will be described in the next subsection), all the DCP histograms extracted in part two and part three are fused together by concatenation, and then input into classifiers for further processing. If KED is chosen, the whole dataset will be grouped into three sets, namely, training set, gallery set and testing set. Firstly, the training set is used to obtain the $s - 1$ projections by KDA, where s is the number of subjects (namely the number of classes). After that, the normal frontal face images and the occluded face images in the training set are applied to learn p kernel principal components which are called the occlusion model in KED. Here, p is set with the default value 10 as in the code exposed by the authors of KED. Please note that the subjects in the training set

[3] Here the DCP histogram under a certain variable combination is denoted by $DCP(BNR, BNC, r_{in}, r_{ex})$. In our method, we extract the following DCP histograms for each face image: DCP(6, 5, 2, 3), DCP(6, 5, 3, 4), DCP(6, 5, 4, 5), DCP(6, 5, 5, 6), DCP(5, 4, 2, 3), DCP(5, 4, 3, 4), DCP(5, 4, 4, 5), DCP(5, 4, 5, 6), DCP(4, 4, 2, 3), DCP(4, 4, 3, 4), DCP(4, 4, 4, 5), DCP(4, 4, 5, 6), DCP(3, 2, 2, 3), DCP(3, 2, 3, 4), DCP(3, 2, 4, 5) and DCP(3, 2, 5, 6). So we get 16 DCP histograms in all for each face image. Please note that we didn't carefully tune these four parameters. According to our experience, the setting of these four parameters will not significantly influence the performance.

cannot exist in the gallery set, as the training set is used only for learning the KDA projections and the occlusion model. Then all gallery samples and occlusion model are projected by KDA projections to get the basic dictionary and extended dictionary, respectively. Finally, we use the sparse representation classifier to classify each probe image in the testing set by minimising the reconstruction residual. For more details of KED, please refer to [36].

2.2 Descriptor selection

The excellent performance of MFD will be demonstrated in Sect. 3. However, it has a noticeable problem—high dimensionality, which consequently leads to high memory cost and high computational cost. Therefore, in this section, we propose a novel optimisation method called discriminant ability-based multi-descriptor selection (DAMS) to reduce the dimensionality of the feature set. The first issue that needs to be addressed is the manner of evaluating the discriminant ability. As we use kernel discriminant analysis in post-processing, keeping the descriptors that can maximise the discriminant ability is a reasonable choice. In discriminant analysis-based methods, the Fisher objective function is commonly used. Thus the following Fisher objective function is initially considered to evaluate the discriminant ability of a set of descriptors:

$$J(W) = \frac{|W^T S_b W|}{|W^T S_w W|}, \tag{1}$$

where S_w and S_b are the within-class scatter matrix and the between-class scatter matrix, respectively.

By maximising $J(W)$, a projective matrix W^* can be found, that is:

$$W^* = \arg\max_W J(W). \tag{2}$$

It can be demonstrated that [47]:

$$J(W^*) = |eigV|, \tag{3}$$

where $eigV$ denotes the eigenvalue matrix of $S_w^{-1}S_b(n)$, which has the form:

$$eigV = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \tag{4}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the eigenvalues of $S_w^{-1}S_b(n)$.

Algorithm 1 Select n descriptors from the whole 65 descriptors according to the discriminative ability of each descriptor subset.

Input: Feature matrix M, where M is a $i*j$ matrix, which includes the i samples and each instance has j dimensions composed by 65 descriptors. Class label L, which is a i dimensional vector.
Output: Index of the selected n descriptors.

Compute matrix $S_w(65)$ and matrix $S_b(65)$ based on all 65 descriptors.
Randomly select n descriptors from 65 descriptors, $D = (d_1, d_2, ..., d_n)$.
Generate $S_w(n)$ and $S_b(n)$ from $S_w(65)$ and $S_b(65)$.
Compute DA based on $S_w(n)$ and $S_b(n)$.
for $k = 1$ to 30 **do** %Iterate for 30 times.
  for $m = 1$ to n **do** %Process $d_1, d_2, \ldots, d_n$ in sequence.
    for 1 to (65-n) **do** %Traverse all the other (65-n) descriptors.
      Replace d_m with the next one of the other (65-n) descriptors.
      Generate $S_w(n)$ and $S_b(n)$ from $S_w(65)$ and $S_b(65)$.
      Compute DA' based on $S_w(n)$ and $S_b(n)$.
      if $DA' > DA$ **then**
        $DA = DA'$
        Update the index of selected descriptors.
      else
        if $random(0,1) < (10^{-(DA-DA')*k^{1.5}} - 0.5)$ **then** %Accept a worse DA at a certain probability.
          $DA = DA'$
          Update the index of selected descriptors.
        end if
      end if
    end for
  end for
  if $k > 20$ and DA doesn't change **then**
    return The index of selected descriptors.
  end if
end for
return The index of selected descriptors.

Based on the settings above, DAMS (Algorithm 1) is designed to select n descriptors from the initial 65 descriptors (16 global descriptors and 49 local descriptors) according to the discriminant ability of each descriptor subset. In DAMS, we formulate an objective function called DA (discriminant ability) according to the specific situation in the experiments. As shown by Eq. (5), DA is a variant of $J(W^*)$, where CFD is the current feature dimension, n is the number of descriptors in the objective subset, and 150 is the approximate mean size of all descriptor features.

$$DA = \frac{lg(\sqrt[3]{J(W^*)})}{(|CFD - 150 * n|/100)^2 + 1} \tag{5}$$

$$= \frac{lg(|\sqrt[3]{eigV}|)}{(|CFD - 150 * n|/100)^2 + 1}. \tag{6}$$

For the numerator of DA, we extract the cube roots of all the diagonal elements of $eigV$ as the range of $|eigV|$ is wide and mostly beyond the range of the double-precision data type. Then the lg of $\left|\sqrt[3]{eigV}\right|$ is calculated to make the numerator and denominator have the same order of magnitude. For the denominator, $(|CFD - 150 * n|/100)^2$ is the penalty factor to balance the selection bias, as the sizes of different descriptor features are different from each other. Without this penalty factor, DA will tend to choose the descriptors that generate high-dimensional features because high-dimensional features usually have stronger discriminant ability, which goes against our original intention—dimensionality reduction. To accelerate the penalty growth and make the numerator and denominator have the same order of magnitude, $|CFD - 150 * n|$ is divided by 100 and squared. From this penalty factor, it can be seen that the farther CFD deviates from $150 * n$ the greater the denominator and the smaller DA. When CFD equals $150 * n$, the denominator degenerates to 1 and the penalty factor loses its efficacy.

Similar to the simulated annealing algorithm, DAMS also choose a worse DA with a certain probability, which can help DAMS to escape from a local optimum. The escape probability used in DAMS is computed as:

$$P(DA, k) = 10^{-(DA - DA') * k^{1.5}} - 0.5, \tag{7}$$

where DA and DA' denote the discriminant abilities obtained at different stages, and k is the current number of iterations. As k increases, $P(DA, k)$ tends to become smaller. When $DA = DA'$, $P(DA, k)$ equals 0.5, which indicates that DA and DA' have the same probability of being accepted.

To accelerate the running of DAMS, two measures are taken. Firstly, the features extracted by each descriptor are processed by PCA to keep 60% of the principal component variances. Secondly, a novel method is presented to accelerate the computing of $S_w(n)$ and $S_b(n)$, where $S_w(n)$ and $S_b(n)$ are the within-class scatter matrix and between-class scatter matrix based on the selected n descriptors, respectively. In DAMS, most of the time is spent on computing $|eigV|$, where the majority of work is calculating $S_w(n)$ and $S_b(n)$ repeatedly. The new idea is to calculate $S_w(65)$ and $S_b(65)$ just for one time. After that, all the $S_w(n)$ and $S_b(n)$ $(1 \le n \le 65)$ are generated directly from $S_w(65)$ and $S_b(65)$.

The matrices S_w and S_b are defined as below:

$$S_b = \sum_{j=1}^{s} N_j(\mu_j - \mu)(\mu_j - \mu)^T, \tag{8}$$

$$S_w = \sum_{j=1}^{s} \sum_{x \in X_j} (x - \mu_j)(x - \mu_j)^T, \tag{9}$$

where s is the class number, $N_j (j = 1, 2, \ldots, s)$ is the instance number of the jth class, $\mu_j (j = 1, 2, \ldots, s)$ is the mean vector of the jth class, and $X_j (j = 1, 2, \ldots, s)$ is the instance set of the jth class.

Let $\mu_j - \mu = \left[a_1^T a_2^T \cdots a_{65}^T\right]^T$, where $a_1, a_2, \ldots, a_{65}$ are the feature vectors extracted by the corresponding descriptors, respectively. We note that $a_1, a_2, \ldots, a_{65}$ are all column vectors and that they may have different size. Then, $(\mu_j - \mu)^T = \left[a_1^T \ a_2^T \ \cdots \ a_{65}^T\right]$.

Thus, $S_b(65)$ can be denoted as

$$S_b(65) = \sum_{j=1}^{s} N_j(\mu_j - \mu)(\mu_j - \mu)^T \tag{10}$$

$$= \sum_{j=1}^{s} N_j \begin{bmatrix} a_1 a_1^T & a_1 a_2^T & \cdots & a_1 a_{65}^T \\ a_2 a_1^T & a_2 a_2^T & \cdots & a_2 a_{65}^T \\ \vdots & \vdots & \ddots & \vdots \\ a_{65} a_1^T & a_{65} a_2^T & \cdots & a_{65} a_{65}^T \end{bmatrix} \tag{11}$$

If we denote $\sum_{j=1}^{s} N_j a_n a_n^T = A_n A_n^T (1 \le n \le 65)$, then

$$S_b(65) = \begin{bmatrix} A_1 A_1^T & A_1 A_2^T & \cdots & A_1 A_{65}^T \\ A_2 A_1^T & A_2 A_2^T & \cdots & A_2 A_{65}^T \\ \vdots & \vdots & \ddots & \vdots \\ A_{65} A_1^T & A_{65} A_2^T & \cdots & A_{65} A_{65}^T \end{bmatrix} \tag{12}$$

It can be seen from (13) that any $S_b(n)$ can be generated from $S_b(65)$ by simply deleting the columns and rows that contain the corresponding descriptors in $S_b(65)$ but out of $S_b(n)$. For example, $S_b(6)$ involves descriptor 3 to descriptor 8, and so it can be generated by deleting the columns and rows that contain the descriptors $1, 2, 9, 10, \ldots, 65$. As a result, we get the $S_b(6)$ as follows:

Fig. 4 Some examples from the CAS-PEAL-R1 dataset

$$S_b(6) = \begin{bmatrix} A_3A_3^T & A_3A_4^T & \cdots & A_3A_8^T \\ A_4A_3^T & A_4A_4^T & \cdots & A_4A_8^T \\ \vdots & \vdots & \ddots & \vdots \\ A_8A_3^T & A_8A_4^T & \cdots & A_8A_8^T \end{bmatrix} \tag{13}$$

It is worth mentioning that by experiment we found that DAMS spent 92.8% of the time ($n = 10$) in calculating $S_w(n)$ and $S_b(n)$ before we use the trick of block matrix operation. Thus this trick is important in making DAMS feasible.

3 Experiments

3.1 Results on the CAS-PEAL-R1 Dataset

In this section, the performance of MDF and SMDF is evaluated on the CAS-PEAL-R1 dataset [39]. The CAS-PEAL-R1 dataset is constructed by the Chinese Academy of Sciences and contains 99,594 images from 1040 subjects (including 595 males and 445 females). In our experiments, we use the following subsets: 'Normal', 'Expression', 'Lighting', 'Accessory', 'Background', 'Distance' and 'Aging', which contain face images from 1040 subjects in total. Some examples from the CAS-PEAL-R1 dataset are shown in Fig. 4.

Following the standard experimental protocol [39], we use the whole 'Normal' subset as the gallery set; it consists of 1040 images from 1040 subjects (one sample per person). For the training set, we randomly select 400 images (100 subjects, 4 samples per person) from the 'Expression' subset, 800 images (200 subjects, 4 samples per person) from the 'Lighting' subset, 80 images (20 subjects, 4 samples per person) from the 'Accessory' subset; and for those subjects who appear in the above mentioned images, we also add their images in the 'Normal' subset into the training set. Excluding the face images used in the training set, the rest of the 'Expression', 'Lighting', 'Accessory', 'Background', 'Distance' and 'Aging' subsets are used to create six probe sets respectively. The face portion of each image is cropped out and normalized to the size of 120*100 pixels. In order to ensure the veracity and reliability of our experimental results, each experiment is repeated ten times. Here the parameter – number of descriptors is set to be 6 in SMDF, as 6 descriptors are already sufficient to achieve a good result.

Table 1 Comparison with SRC-based methods on the CAS-PEAL-R1 dataset

Method	Mean accuracy (%) ± std. dev.		
	Accessory	Lighting	Expression
MLBP + SRC	72.9 ± 0.6	17.3 ± 0.7	98.2 ± 0.4
MLBP + ESRC	87.1 ± 1.0	82.1 ± 0.4	99.7 ± 0.1
MLBP + KDA+SRC	80.8 ± 1.9	82.7 ± 0.6	99.7 ± 0.1
MLBP + KDA+ESRC	80.9 ± 1.9	83.0 ± 0.6	99.7 ± 0.1
MLBP + KED	91.0 ± 0.6	83.1 ± 0.5	99.7 ± 0.1
MDF + KED*	**97.5 ± 0.3**	**85.5 ± 1.6**	**99.7 ± 0.1**
SMDF(6) + KED*	**95.2 ± 0.5**	**85.0 ± 1.3**	**99.4 ± 0.2**

The proposed methods are highlighted with asterisks and their results are marked in bold

3.1.1 Comparison with SRC-based methods

KED is an SRC-based method. To demonstrate the effectiveness of combining MDF/SMDF(6) with KED, we compare MDF/SMDF(6) + KED with other SRC-based methods, including SRC [28], ESRC [37], KDA + SRC, KDA + ESRC and KED [36]. For initial features, we use the same Multiscale LBP (MLBP) features as in [36] so as to maximise the performance of these baseline methods. Table 1 shows the results of different methods on three subsets. According to the results, we can observe the following:

1. All methods perform well on the Expression probe set. This is because SRC-based methods are robust to local variances such as expression and local occlusion [36].
2. MDF + KED and SMDF + KED significantly outperform other methods on the Accessory probe set; KED performs better than other methods, but it is inferior to MDF + KED and SMDF + KED. The Accessory probe set contains a large number of cases of contiguous occlusion. SRC, ESRC, KDA + SRC and KDA + ESRC fail to handle these contiguous occlusions, so they perform poorly on this probe set.
3. All methods perform relatively poorly on the Lighting probe set, but MDF + KED and SMDF + KED are still better than other SRC-based methods by at least 2.4% and 1.7%, respectively. In this case, even though the standard deviation of MDF+KED reaches 1.6, it is still acceptable.

Table 2 Comparison with other descriptors on the CAS-PEAL-R1 dataset

Method	Recognition accuracy (%)					
	Accessory	Lighting	Expression	Time	Background	Distance
LBP	91.82	46.90	94.27	100.00	99.46	44.60
LTP	91.77	47.17	94.39	100.00	99.46	44.68
LPQ	92.39	57.16	93.95	100.00	99.28	44.76
POEM	92.39	54.66	95.54	100.00	99.46	42.52
LGXP	91.33	63.26	94.97	100.00	99.28	22.91
MsLBP	92.04	47.75	95.16	100.00	99.46	44.88
MsTLBP	92.74	48.06	95.41	100.00	99.46	45.48
MsDLBP	90.63	48.11	92.42	100.00	99.28	37.03
DCP	92.82	50.25	96.11	100.00	99.10	51.30
MDF + KED*	**97.47**	**85.49**	**99.67**	**100.00**	**99.94**	**100.00**
SMDF(6) + KED*	**95.66**	**85.09**	**99.73**	**100.00**	**99.39**	**100.00**

The proposed methods are highlighted with asterisks and their results are marked in bold

Table 3 Comparison with state-of-the-art methods on the CAS-PEAL-R1 dataset

Method	Mean recognition accuracy (%)					
	Accessory	Lighting	Expression	Time	Background	Distance
SSEC TIP13'	66.6	17.4	74.5	51.9	66.8	84.2
RRC TIP13'	84.2	29.3	94.0	96.7	95.6	97.9
SLF-RKR TNNLS13'	90.9	28.8	99.6	98.5	99.9	99.7
MOST TIP14'	80.4	82.4	98.2	97.9	99.0	99.8
KED TNNLS16'	91.0	83.1	99.7	99.7	99.9	99.9
DCP TPAMI16'	92.8	50.3	96.1	100.0	99.1	51.3
MDF + KED*	**97.5**	**85.5**	**99.7**	**100.0**	**99.9**	**100.0**
SMDF(6) + KED*	**95.7**	**85.1**	**99.7**	**100.0**	**99.4**	**100.0**

The proposed methods are highlighted with asterisks and their results are marked in bold

3.1.2 Comparison with other descriptors

We also compare the proposed method with other descriptor-based methods. They are LBP [17], LTP [48], LPQ [49], POEM [50], local gabor XOR patterns (LGXP) [7], multiscale LBP [51], multiscale tLBP (MsTLBP) [52], multiscale dLBP (MsDLBP) [52] and DCP [15]. To maximise the performance of these baseline descriptors, we carefully choose the parameters and the distance functions (Chi-squared or histogram intersection) for each of them. The final results are reported in Table 2 with the parameters and distance functions that can maximise the average accuracy on all subsets. According to the results, we can observe the following:

1. MDF + KED has the best identification rates on all six probe sets, which demonstrates the superiority of the proposed method.
2. MDF + KED and SMDF + KED perform much better than DCP, which demonstrates the effectiveness of the fused descriptor features extracted by MDF and SMDF.
3. The results on the lighting and distance probe sets indicate that the baseline methods misclassify a large proportion of the probe images on these two probe sets. An explanation is that the images from these two probe sets are significantly overlapped in the feature space, and the baseline methods have insufficient features that have strong discriminant ability.

3.1.3 Comparison with state-of-the-art methods

Finally we compare the proposed MDF+KED and SMDF+KED with the state-of-the-art methods, including SSEC [33], RRC [34], SLF-RKR [35], MOST [53], KED and DCP. For the parameters of SSEC, they are set as in [33]: $\lambda_E = 2$, $\lambda_V = 0$, $\kappa = 0.3$, and $T = 5$. For the parameters of RRC, we followed the settings of [34]: $\mu = (\varsigma/\delta)$, $\varsigma = 8$, and $\tau = 0.8$. For SLF-RKR, we set $S = 0$, $P_0 = 5$, and $Q_0 = 4$ as presented in [35]. For the settings of KED and DCP, we followed [15, 36], respectively. The comparative results are shown in Table 3, from which we can observe the following:

1. Compared with the state-of-the-art methods, MDF+KED and SMDF + KED still have the best iden-

Fig. 5 Some examples from the LFW dataset

Table 4 Comparison with state-of-the-art methods on the LFW dataset

	Method	Recognition accuracy
Without outside training data	PCA600	56.9
	KDA+1NN	40.0
	KDA+SRC	89.2
	KED	89.2
	MDF*	**94.3**
	SMDF(11)*	**91.7**
With outside training data	COTS-s1 [54]	56.7
	COTS-s1 + s4 [54]	66.5
	DeepFace [10]	64.9
	WST Fusion [55]	82.5
	DeepID2+ [56]	95.0

The proposed methods are highlighted with asterisks and their results are marked in bold

tification rates on all six probe sets, which again demonstrates the good performance of the proposed methods.
2. Most methods perform well on Expression, Time and Background probe sets. However, they perform poorly on the Accessory probe set. MDF + KED achieves a better result on the Accessory probe set than other methods.
3. DCP has good performance on Accessory, Expression, Time and Background probe sets, but cannot cope well with the Lighting and Distance probe sets. KED has excellent results on all six probe sets except the Accessory probe set.

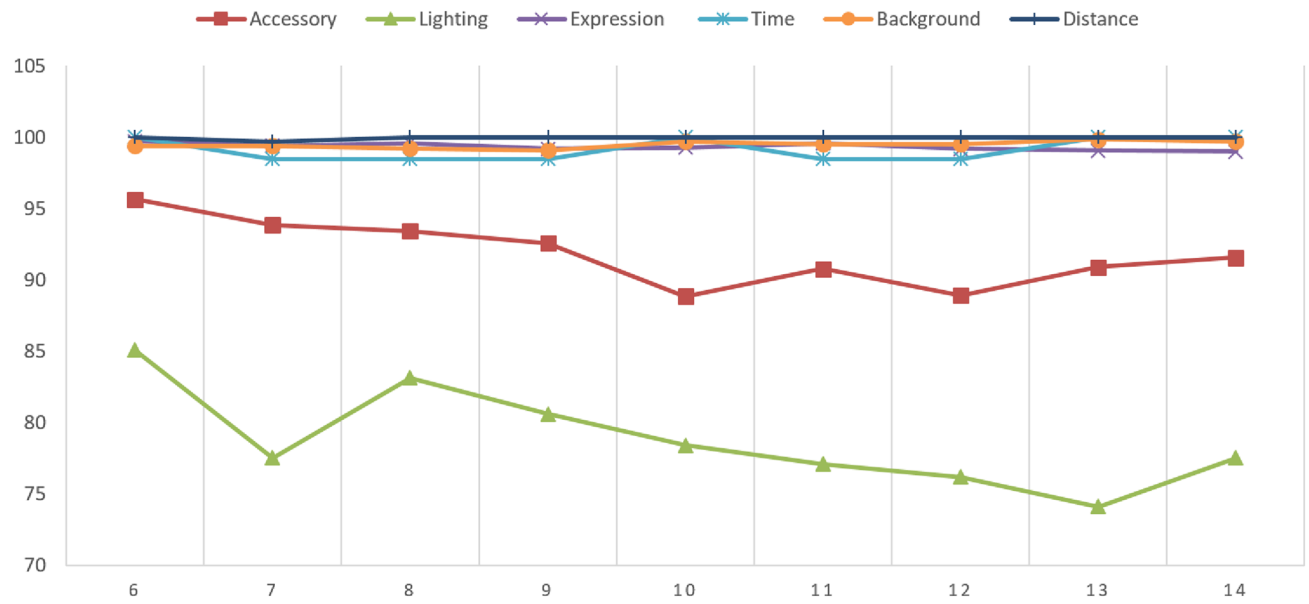
3.2 Results on the LFW dataset

The LFW face dataset [1] consists of more than 13,000 facial images of 5749 subjects downloaded form the web, and has been created for research on unconstrained face recognition. The facial images in the LFW dataset have dramatic variations of illumination, occlusion, pose and expression; the only constraint is that all these faces were captured by the Viola-Jones face detector. Currently, there are four different versions of LFW, including the original version and three different types of "aligned" versions. In the following experiment we use the version called "LFW-a". Some original facial images from the LWF dataset are shown in Fig. 5. As preprocessing, all the images in LFWa are normalised to 120*100 pixels and processed by affine transform based on three fiducial marks (left eye centre, right eye centre and mouth centre) obtained by the SDM algorithm. We use the mean value of landmarks 20–25 to get the position of the left eye centre, use the mean value of landmarks 26–31 to get the position of the right eye centre, and use the mean value of landmarks 32, 38 and 44 to 49 to get the position of the mouth centre (see Fig. 2). In the processing of the affine transform, all the face images are aligned with the left eye centre, right eye centre and mouth centre mapped to (29, 42), (75, 42) and (53, 96), respectively.

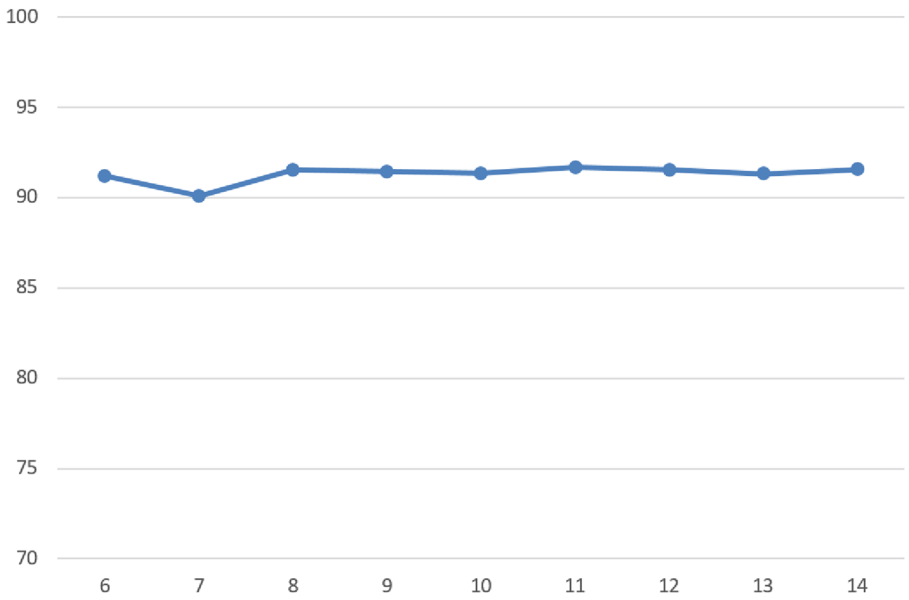
To demonstrate the effectiveness of MDF and SMDF, we compare the proposed methods with PCA600 (reduce to 600 dimensions by PCA), KDA + 1NN, KDA + SRC and KED. Different from the settings used on the CAS-PEAL-R1 dataset, the Cosine KNN classifier is used for the proposed MDF and SMDF in this section. We explored a number of classifiers and Cosine KNN shows the best performance in this case. Following the experimental protocol in [35, 36], a subset of LFW is used in the experiments, which contains 5425 images of 311 subjects with no less than six samples per subject. The parameters of these methods are the same as the settings in Sect. 3.1. To avoid overfitting, fivefold cross-validation is applied with all of the above methods.

Additionally, we also include comparable results on the same data by deep learning based methods—COTS-s1 [54], COTS-s1+s4 [54], WST Fusion [55], DeepFace [10], and DeepID2+ [56]. Worthy of noting is that DeepFace and DeepID2+ are two representative deep learning-based methods for face recognition. The experimental results of different methods are presented in Table 4, arranged into two categories, according to the experimental routine of LFW, namely the methods with outside training data and the methods without outside training data.

1. Most methods do not achieve good results on the LFW dataset because of its unconstrained and dramatic variations.
2. MDF and SMDF significantly outperform the other methods without outside data, while the identification accuracy of MDF is slightly higher than SMDF.
3. Compared with deep learning-based methods, MDF and SMDF are still competitive. From the perspective of identification accuracy, MDF and SMDF are better than DeepFace and WST Fusion, but a little worse than DeepID2+.



(a) Accuracy VS number of descriptors on CAS-PEAL-R1



(b) Accuracy VS number of descriptors on LFW

Fig. 6 The relationship between the identification accuracy and the number of descriptors on CAS-PEAL-R1 and LFW dataset

Table 5 Runtime evaluation on LFW dataset

	PCA600	KDA + 1NN	KDA + SRC	KED	MDF	SMDF(14)	SMDF(8)	SMDF(6)
Initial feature dimension	17,110	17,110	17,110	17,110	247,808	51,200	52,224	53,248
Training time (h)	0.7	0.9	0.9	1	11.1	2.7	2.7	2.7
Classification time per instance (ms)	8	21	20	20	187	40	41	41
Max memory usage (GB)	2.6	2.3	2.3	2.4	22.0	6.0	6.2	6.3
Accuracy on LFW (%)	56.9	40.0	89.2	89.2	94.3	91.6	91.5	91.2

3.3 Stability and runtime evaluation

In this section, the stability of DAMS and the runtimes of the proposed methods are discussed. Firstly, we explore the relationship between the identification accuracy and the number of descriptors. The number of descriptors is an important parameter in DAMS, which determines the number of target descriptors, leading to different initial feature dimension and different identification accuracy. Therefore, the following experiments were conducted on the CAS-PEAL-R1 and LFW datasets based on the same experimental settings asin Sects. 3.1 and 3.2. In this process, we run DAMS for 5 times based on different numbers of descriptors and select the descriptor subsets that can maximise the objective function. As shown in Fig. 6, the identification accuracies on LFW and most subsets of CAS-PEAL-R1 vary little as the number of descriptors changes from 6 to 14. But the accuracies on Accessory and Lighting subsets show a slow downward trend. A reasonable explanation is that the proposed objective function—DA specifies the number of descriptors, but it does not specify the target dimension (it only uses a penalty factor to balance the selection bias). Whereas the global descriptors have higher dimensions than the local descriptors, they are more helpful in enhancing the value of DA. So DAMS tends to choose more global descriptors rather than local descriptors as the number of descriptors increases. However, local features can cope better with the Accessory and Lighting subsets, which include different type of occlusion and illumination variations. In summary, the number of descriptors is not a sensitive parameter, but it is worth selecting a value carefully when handling some specific situations like occlusion and illumination variations.

Using a single thread with 3.47 GHz CPU (Intel Xeon X5690), we conducted experiments on the LFW dataset and recorded the runtime and relevant details of the proposed methods and some other methods we implemented. Results are shown in Table 5. PCA600, KDA+1NN, KDA+SRC and KED have lower requirement on memory usage and runtime on training and classification, but they have much lower recognition accuracies. For example their recognition accuracies are all lower than 90%, while the proposed MDF and SMDF have accuracies of 94.3% and 91.6%, respectively. Compared with MDF, SMDF has a much smaller feature set, which is only approximately one-fifth of the feature set of MDF. The reduction in feature dimension leads to lower computational cost and memory cost. The training time decreases from 11.1 to 2.7 h, while the classification time drops from 187 to 40 ms. Another important change is in the maximum memory cost, which is reduced to only approximately 6 GB in SMDF from the 22 GB in MDF. This enables a typical modern computer with 8 GB memory to run the proposed face identification algorithm. As a compromise, we lose 2.9% $\pm$ 0.2% accuracy, but even so, the performance of SMDF is still better than many of the state-of-the-art methods, as illustrated in Table 4.

4 Conclusion

To fully utilise the discriminant information and improve the discriminative ability of features, in this paper we propose a high-performance face image representation method—MDF, by which we achieved higher identification accuracy than the state-of-the-art methods. Further still, we propose a novel optimisation method, DAMS, which reduces the computational cost and the memory cost of MDF. Compared with MDF, the DAMS-optimised face representation, SMDF, has much smaller feature dimension, resulting in a much lower configuration requirement. However, SMDF still achieves excellent performance compared with other state-of-the-art methods.

References

1. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Tech. Rep. pp 07–49
2. Sun Y, Liang D, Wang X, Tang X (2015) Deepid3: Face recognition with very deep neural networks. arXiv preprint arXiv :1502.00873
3. Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: a unified embedding for face recognition and clustering. In: pp 815–823
4. Liu J, Deng Y, Bai T, Wei Z, Huang C (2015) Targeting ultimate accuracy: face recognition via deep embedding, pp 06–24. arXiv :1506.07310[cs]
5. Kumar N, Berg A C, Belhumeur P N, Nayar S K (2009) Attribute and simile classifiers for face verification. In: 2009 IEEE 12th international conference on computer vision, Sep., pp 365–372
6. Learned-Miller E, Huang GB, RoyChowdhury A, Li H, Hua G (2016) Labeled faces in the wild: a survey. In: Kawulok M, Celebi ME, Smolka B (eds) Advances in face detection and facial image analysis. Springer International Publishing, Berlin, pp 189–248. https://doi.org/10.1007/978-3-319-25958-1_8
7. Xie S, Shan S, Chen X, Chen J (2010) Fusing local patterns of gabor magnitude and phase for face recognition. IEEE Trans Image Process 19(5):1349–1361
8. Wei X, Wang H, Guo G, Wan H (2015) Multiplex image representation for enhanced recognition. Int J Mach Learn Cybern 9:1–10
9. Chan CH, Tahir MA, Kittler J, Pietikäinen M (2013) Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. IEEE Trans Pattern Anal Mach Intell 35(5):1164–1177
10. Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: closing the gap to human-level performance in face verification. In: pp 1701–1708
11. Sun Y, Wang X, Tang X (2014) Deep learning face representation from predicting 10,000 classes. In: pp 1891–1898
12. Ding C, Tao D (2017) Trunk-Branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans Pattern Anal Mach Intell PP(99):1–1

13. Huang GB, Learned-Miller E (2014) Labeled faces in the wild: Updates and new reporting procedures. Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep, pp 14–003
14. Huang KK, Dai DQ, Ren CX, Yu YF, Lai ZR (2017) Fusing landmark-based features at kernel level for face recognition. Pattern Recognit 63:406–415
15. Ding C, Choi J, Tao D, Davis LS (2016) Multi-directional multi-level dual-cross patterns for robust face recognition. IEEE Trans Pattern Anal Mach Intell 38(3):518–531
16. Visual descriptor (2017) Page Version ID: 788982618. [Online]. https://en.wikipedia.org/w/index.php?title=Visual_descriptor&oldid=788982618
17. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell 28(12):2037–2041
18. Qi X, Xiao R, Li CG, Qiao Y, Guo J, Tang X (2014) Pairwise rotation invariant co-occurrence local binary pattern. IEEE Trans Pattern Anal Mach Intell 36(11):2199–2213
19. Kim J, Yu S, Kim D, Toh K-A, Lee S (2017) An adaptive local binary pattern for 3d hand tracking. Pattern Recognit 61(Supplement C):139–152. [Online]. http://www.sciencedirect.com/science/article/pii/S0031320316301972
20. Karczmarek P, Kiersztyn A, Pedrycz W, Dolecki M (2017) An application of chain code-based local descriptor and its extension to face recognition. Pattern Recognit 65:26–34
21. Zhen X, Zheng F, Shao L, Cao X, Xu D (2017) Supervised local descriptor learning for human action recognition. IEEE Trans Multimed 19(9):2056–2065
22. Lan R, Zhou Y, Tang YY (2016) Quaternionic local ranking binary pattern: a local descriptor of color images. IEEE Trans Image Process 25(2):566–579
23. Yan P, Liang D, Tang J, Zhu M (2016) Local feature descriptor using entropy rate. Neurocomputing 194:157–167
24. Mangai UG, Samanta S, Das S, Chowdhury PR (2010) A survey of decision fusion and feature fusion strategies for pattern classification. IETE Tech Rev 27(4):293–307. [Online]. http://www.tandfonline.com/doi/abs/10.4103/0256-4602.64604
25. Nikan S, Ahmadi M (2014) Local gradient-based illumination invariant face recognition using local phase quantisation and multi-resolution local binary pattern fusion. IET Image Process 9(1):12–21
26. Gao Z, Ding L, Xiong C, Huang B (2014) A robust face recognition method using multiple features fusion and linear regression. Wuhan Univ J Nat Sci 19(4):323–327
27. Ruggieri S (2002) Efficient C4.5 [classification algorithm]. IEEE Trans Knowl Data Eng 14(2):438–444
28. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
29. Weinberger KQ, Saul LK (2009) Distance metric learning for large margin nearest neighbor classification. J Mach Learn Res 10:207–244 [Online]. http://www.jmlr.org/papers/v10/weinberger09a.html
30. Heisele B, Ho P, Poggio T (2001) Face recognition with support vector machines: global versus component-based approach. In: Proceedings of eighth IEEE international conference on computer vision, vol 2. ICCV 2001, pp 688–694
31. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113
32. Sebe N, Lew MS, Cohen I, Garg A, Huang TS (2002) Emotion recognition using a Cauchy Naive Bayes classifier. In: Object recognition supported by user interaction for service robots, vol 1, pp 17–20
33. Li XX, Dai DQ, Zhang XF, Ren CX (2013) Structured sparse error coding for face recognition with occlusion. IEEE Trans Image Process 22(5):1889–1900
34. Yang M, Zhang L, Yang J, Zhang D (2013) Regularized robust coding for face recognition. IEEE Trans Image Process 22(5):1753–1766
35. Yang M, Zhang L, Shiu SCK, Zhang D (2013) Robust kernel representation with statistical local features for face recognition. IEEE Trans Neural Netw Learn Syst 24(6):900–912
36. Huang KK, Dai DQ, Ren CX, Lai ZR (2016) Learning kernel extended dictionary for face recognition. IEEE Trans Neural Netw Learn Syst PP(99):1–13
37. Deng W, Hu J, Guo J (2012) Extended SRC: undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell 34(9):1864–1870
38. Deng W, Hu J, Guo J (2013) In defense of sparsity based face recognition. In The IEEE conference on computer vision and pattern recognition (CVPR)
39. Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D (2008) The CAS-PEAL large-scale chinese face database and baseline evaluations. IEEE Trans Syst Man Cybern Part A Syst Hum 38(1):149–161
40. Xiong X, De la Torre F (2013) Supervised descent method and its applications to face alignment. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 532–539
41. Gross R, Matthews I, Cohn J, Kanade T, Baker S (2010) Multi-pie. Image Vis Comput 28(5):807–813
42. Saragih J (2011) Principal regression analysis. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2881–2888
43. Bartlett MS, Littlewort G, Frank MG, Lainscsek C, Fasel IR, Movellan JR (2006) Automatic recognition of facial actions in spontaneous expressions. J Multimed 1(6):22–35
44. Tan H, Yang B, Ma Z (2013) Face recognition based on the fusion of global and local hog features of face images. IET Comput Vis 8(3):224–234
45. Fierro-Radilla AN, Nakano-Miyatake M, Perez-Meana H, Cedillo-Hernandez M, Garcia-Ugalde F (2013) An efficient color descriptor based on global and local color features for image retrieval. In: IEEE 2013 10th international conference on electrical engineering, computing science and automatic control (CCE), pp 233–238
46. Shabanzade M, Zahedi M, Aghvami SA (2011) Combination of local descriptors and global features for leaf recognition. Signal Image Process 2(3):23
47. Swets DL, Weng JJ (1996) Using discriminant eigenfeatures for image retrieval. IEEE Trans Pattern Anal Mach Intell 18(8):831–836
48. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650
49. Ahonen T, Rahtu E, Ojansivu V, Heikkila J (2008) Recognition of blurred faces using local phase quantization. In 2008 19th international conference on pattern recognition, vol 12, pp 1–4
50. Vu NS, Caplier A (2012) Enhanced patterns of oriented edge magnitudes for face recognition and image matching. IEEE Trans Image Process 21(3):1352–1365
51. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
52. Trefnỳ J, Matas J (2010) Extended set of local binary patterns for rapid object detection. In: Computer vision winter workshop, pp 1–7
53. Ren CX, Dai DQ, Li XX, Lai ZR (2014) Band-reweighed gabor kernel embedding for face image representation and recognition. IEEE Trans Image Process 23(2):725–740

54. Best-Rowden L, Han H, Otto C, Klare BF, Jain AK (2014) Unconstrained face recognition: identifying a person of interest from a media collection. IEEE Trans Inf Forensics Secur 9(12):2144–2157
55. Taigman Y, Yang M, Ranzato M, Wolf L (2015) Web-scale training for face identification. In: pp 2746–2754. [Online]. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Taigman_Web-Scale_Training_for_2015_CVPR_paper.html
56. Sun Y, Wang X, Tang X (2015) Deeply learned face representations are sparse, selective, and robust. In: pp 2892–2900. [Online]. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Sun_Deeply_Learned_Face_2015_CVPR_paper.html