Signal Processing 144 (2018) 296–305

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Feature fusion with covariance matrix regularization in face recognition

Ze Lu*, Xudong Jiang, Alex Kot

Nanyang Technological University, 50 Nanyang Drive, 637553, Singapore

ARTICLE INFO

Article history:
Received 24 August 2017
Accepted 22 October 2017
Available online 5 November 2017

Keywords:
Feature fusion
CNN
Overfitting
Regularization
Face recognition

ABSTRACT

The fusion of multiple features is important for achieving state-of-the-art face recognition results. This has been proven in both traditional and deep learning approaches. Existing feature fusion methods either reduce the dimensionality of each feature first and then concatenate all low-dimensional feature vectors, named as DR-Cat, or the vice versa, named as Cat-DR. However, DR-Cat ignores the correlation information between different features which is useful for classification. In Cat-DR, on the other hand, the correlation information estimated from the training data may not be reliable especially when the number of training samples is limited. We propose a covariance matrix regularization (CMR) technique to solve problems of DR-Cat and Cat-DR. It works by assigning weights to cross-feature covariances in the covariance matrix of training data. Thus the feature correlation estimated from training data is regularized before being used to train the feature fusion model. The proposed CMR is applied to 4 feature fusion schemes: fusion of pixel values from 3 color channels, fusion of LBP features from 3 color channels, fusion of pixel values and LBP features from a single color channel, and fusion of CNN features extracted by 2 deep models. Extensive experiments of face recognition and verification are conducted on databases including MultiPIE, Georgia Tech, AR and LFW. Results show that the proposed CMR technique significantly and consistently outperforms the best single feature, DR-Cat and Cat-DR.

1. Introduction

Face recognition has been a very active research area due to its increasing security demands, commercial applications and law enforcement applications [1–6]. It is often the case in face recognition that no single feature is rich enough to capture all of the available information [7]. The robust face recognition requires multiple feature sets to be taken into account [8], which can be features of different color channels [9–12], different types of features [8,13,14] and features extracted by different deep models [15–17]. Recently, Convolutional Neural Networks (CNN) provides an effective tool for feature learning in face recognition and very promising results have been obtained as in [18,19]. The pre-trained VGG-Face model [18] was learned from a large face dataset containing 2.6M web images of 2622 celebrities and public figures. It is widely used as a feature extractor for classifying face images as in [20–22]. Different from the architecture of VGG-Face, ResNet in [19] consists of residual modules which conduct additive merging of signals. The authors in [19] argue that residual connections are inherently important for training very deep architectures. It is natural to study the combination of VGG-Face with ResNet, which would allow two models to reap the benefits of each other. Thus we train a ResNet-like CNN model using images from the recently released CASIA-WebFace dataset [23] and combine it with the pre-trained VGG-Face model by feature fusion.

Feature fusion often results in very high dimensionality. For example, multi-scale descriptors in [24] are densely extracted from dense landmarks and concatenated together to form a 100K-dimensional feature vector. The high dimensionality of feature vectors imposes great burdens on the robust face recognition task. Therefore, dimensionality reduction is a critical module of feature fusion. Existing feature fusion methods can be generally classified into two categories: DR-Cat and Cat-DR. DR-Cat applies dimensionality reduction to each feature before the concatenation of multiple features and Cat-DR does vice versa. Choi and et al. [11] use DR-Cat to reduce the dimension of each color local texture feature separately before concatenating all low-dimensional features in the column order. Tan and et.al [13] use PCA to reduce the dimensionality of Gabor wavelets and LBP prior to fusing them by averaging their similarity scores (same as DR-Cat). DR-Cat is also used in [12,25–27]. By reducing the dimensionality of each feature separately before concatenating them together, DR-Cat ignores the correlation

* Corresponding author.
E-mail addresses: zlu008@ntu.edu.sg (Z. Lu), exdjiang@ntu.edu.sg (X. Jiang), eackot@ntu.edu.sg (A. Kot).

https://doi.org/10.1016/j.sigpro.2017.10.024
0165-1684/

information between different features. But the correlation information plays an important role in the process of feature fusion. In order to utilize the correlation information, Yang and et.al [28] employ Cat-DR to concatenate three color components into one pattern vector first and then perform PCA or EFM on the concatenated pattern vector. Cat-DR is also used in [24] to fuse multi-scale descriptors centered at dense facial landmarks. The dimension of the concatenated feature is reduced by PCA and LDA. Multiple deep ConvNets are used in [15] to learn face features from images of various scales, where Cat-DR is employed by applying PCA to the concatenation of multiple features. In the case of perfect training data, Cat-DR utilizing the correlation information usually achieves better performance than DR-Cat. However, in practice, the limited training data may result in unreliable estimates of cross-feature correlations. This often leads to overfitting and performance degradation in Cat-DR.

To solve problems in feature fusion methods of DR-Cat and Cat-DR, we propose a covariance matrix regularization (CMR) technique. Instead of modifying eigenvalues of covariance matrices as in conventional regularization techniques [29–33], CMR works by regularizing the off-diagonal cross-feature covariances in the covariance matrix of training data. Thus the trace of covariance matrices remains unchanged and the feature correlation estimated from the training data is suppressed before being used to train the feature fusion model. In this way, the obtained model does not adapt too much to the estimated correlation and hence the overfitting is reduced. In the experimental part conducted on four public face databases including MultiPIE, GT, AR and LFW, we first show that our proposed ResNetShort model achieves state-of-the-art face verification performance on LFW. After that, we vary the value of weights in CMR to show how it solves the problem of overfitting and improves the face recognition performance. Then, we study the relationship between the optimal value of weights in CMR and the number of training images per subject. Finally, we compare the performance of CMR against the best single feature, DR-Cat and Cat-DR by fusing features of multiple color channels, multiple types of features, and features extracted by multiple deep models.

2. Feature fusion in face recognition

2.1. Feature fusion schemes

Face recognition is an area that is well-suited for the fusion of multiple descriptors due to its inherent complexity and need for fine distinctions [8]. Multiple descriptors can be features extracted from different color channels. Y, I, Q components possess the property of decorrelation, which helps reduce redundancy and is an important property in pattern classifier design. Thus features extracted from Y, I, Q color channels are fused in [9]. Similarly, R, Q, Cr features are fused in [10,11] and Z, R, G features are fused in [12]. Furthermore, multiple descriptors can be different types of features. Authors in [8,13] combine Gabor wavelets and LBP to achieve considerably better performance than either alone. The two features are complimentary in the sense that LBP captures small appearance details while Gabor wavelets encode facial shape over a broader range of scales. Fourier features, Gabor wavelets are combined in [14] to achieve better performance for face recognition. Global Fourier features describe the general characteristics of the holistic face and they are often used for coarse representation. Differently, local Gabor features reflect and encode more detailed variations within some local facial regions. Moreover, multiple features may be extracted using different deep models. Authors in [15] train 60 ConvNets, each of which extracts two 160-dimensional DeepID vectors from 60 face patches with ten regions, three scales, and RGB or gray channels. Combing 60 different deep models increases the face verification accuracy by 5.27% over the best single model. The deep learning structure proposed in [16] is composed of a set of elaborately designed CNN models, which extract complementary facial features from multimodal facial data.

To investigate the effectiveness of the proposed feature fusion method for face recognition, this paper explores 4 different feature fusion schemes: (1), fusion of pixel values in 3 color channels *R, G, B*; (2), fusion of LBP features in 3 color channels *R, G, B*; (3), fusion of pixel values and LBP features of a single color channel *R*; (4), fusion of CNN features extracted by 2 deep models. Many recent face recognition works conduct experiments on pixel values to evaluate the face recognition performance of their methods [34–37]. LBP has been proven to be highly discriminative for face recognition [24,38]. Thus these two features are used for the task of fusing features of different color channels *R, G, B* and the task of fusing different types of features in channel *R*. As *R* channel has been shown to perform better than other intensity images including Gray for face retrieval [11,34], we take the *R* channel as an example channel for the fusion of different types of features. For the fusion of multiple deep learning features, we utilize the pre-trained VGG-Face model and propose a new deep model, ResNet-Short, presented in the following section.

2.2. Deep learning feature fusion: VGG-Face and ResNetShort

Convolutional Neural Networks have significantly improved the state of the arts in face recognition [39]. VGG-Face is a deep neural network proposed by Simonyan et al. in [18]. This network is characterized by using 3×3 convolutional layers stacked on top of each other in increasing depth. The architecture of VGG-Face comprises 21 layers, which consist of 13 convolutional layers, 5 maxpooling layers and 3 fully connected layers. The first two fully connected layers are 4096 dimensional and the dimension of the last fully connected layer depends upon the loss functions used for optimisation. The pre-trained VGG-Face model was learned from a large face dataset (see Fig. 1 for sample images) containing 2.6M images of 2622 celebrities and public figures. Faces are detected using the method described in [40] and a 2D similarity transformation is applied to map the face to a canonical position. VGG-Face is first trained as a multi-class classification problem by minimizing the softmax loss and then fine-tuned by the recently proposed triplet loss [41]. The pre-trained VGG-face model has been widely used by researchers to extract CNN features from face images as in [20–22].

Unlike traditional sequential network architectures such as VGG, ResNet consists of “network-in-network” modules. First introduced by He et al. in [19], ResNet has become a seminal work, demonstrating that the degradation problem of deep networks can be solved through the use of residual modules. ResNet layers are formulated as learning residual functions with reference to the layer inputs. By referring to the CNN model used in [42] and residential modules, we propose a model as shown in Fig. 2 and name it ResNetShort. The size of filters in convolution layers is 3×3 with stride 1, followed by PReLU [43] non-linear units. The max-pooling grid is 2×2 and the stride is 2. The number of feature maps in convolutional layers or the dimension of fully connected layers is indicated by the number on top of each layer. ‘$\times h$’ represents a residual module that repeats for h times. Joint supervision of softmax loss and center loss [42] is adopted. The value of λ, which is used for balancing the softmax and center loss functions, is set as 0.005.

The recently released CASIA-WebFace [23] database is used to train the ResNetShort model. CASIA-WebFace contains 494,414 images of 10,575 subjects. According to [44], adding the individuals with only a few instances do not help to improve the recognition performance. Indeed, these individuals will harm the systems performance. Thus the 10,575 subjects are ranked in the descent or-

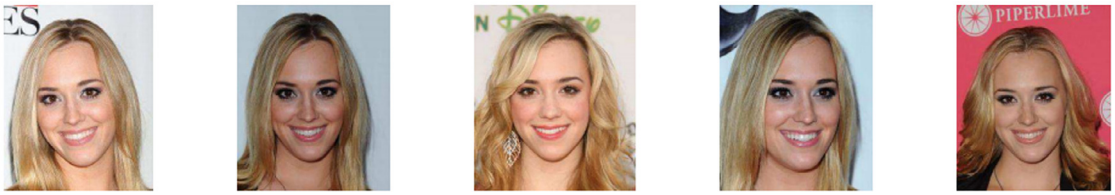

Fig. 1. Sample face images from VGG Face database.

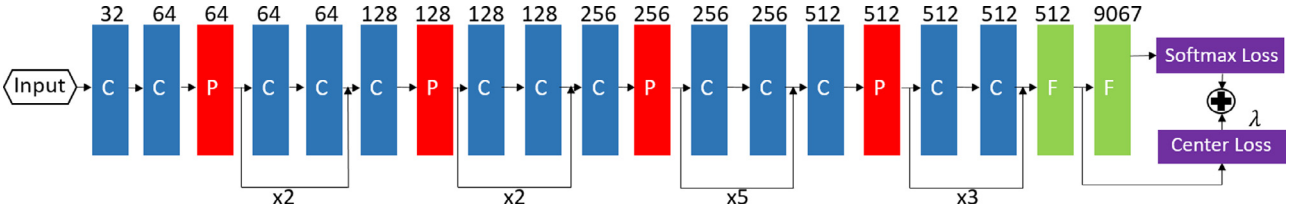

Fig. 2. The ResNetShort architecture, where C, P, and F indicate convolutional, max pooling, and fully connected layers, respectively.

Fig. 3. Normalized face images from CASIA-WebFace database.

Table 1
Comparison between the pre-trained VGG-Face model and our trained ResnetShort model, CONV and FC indicate convolutional and fully connected layers, respectively.

Model	VGG-Face	ResNetShort
Training data	VGG Face	CASIA-WebFace
Face alignment	vanilla DPM [40]	TCDCN [45]
Input size	$224 \times 224 \times 3$	$112 \times 96 \times 3$
Architecture	CONV + FC	Residual modules
Non-linear units	ReLU	PReLu
Feature size	4096	512
Supervision signals	softmax+triplet loss	softmax+center loss

der by the number of their images contained in the database. The 434,793 images of the top 9067 subjects, which contain at least 14 images per subject, compose the training set. The remaining images of the rest 1508 subjects are discarded. Face images are normalized to 112×96 pixels with an affine transformation according to the coordinates of five sparse facial points, i.e., both eye centers, the nose tip, and both mouth corners. Sample images after the affine transformation are shown in Fig. 3. We employ an off-the-shelf face alignment tool [45] for facial point detection and double the size of the training set by flipping all training images horizontally. The open-source deep learning toolkit Caffe [46] is utilized to train the deep model. During training, the batch size is set to 256. The initial learning rate for all learning layers is set to 0.1, and is divided by 10 after 16,000 iterations, and is then divided by 10 after 8000 iterations to the final rate of 0.001. The total number of iterations is 28000.

Both the pre-trained VGG-Face model and our proposed and trained ResNetShort model achieve state-of-the-art face verification performance on challenging face datasets (refer to Section 5.1). A comprehensive comparison between these two deep models is given on Table 1. From which we can observe that, these two models are trained from different face images by optimizing different loss functions through different deep architectures. This makes the learned discriminative information contained in VGG-Face features and ResNetShort features mutually complementary to each other. Therefore, we combine these two CNN models by feature fusion to effectively make use of their discriminative information. Feature fusion methods for face recognition are discussed in following sections.

3. Feature fusion with dimensionality reduction

Fusing multiple feature sets has many successful applications in face recognition. However, the fusion of multiple features inevitably causes the problem of high dimensionality. It is well known that high dimensionality degrades the classification performance (curse of dimensionality) [47,48]. Thus, dimension reduction becomes an integrated part of feature fusion. PCA [49] is commonly used as a benchmark for the evaluation of the performance in FR algorithms [50] and it may significantly enhance the recognition accuracy [51,52]. Plenty of color face recognition methods adopt the Enhanced Fisher Model (EFM) [35,36,53]. Therefore, PCA and EFM are used in this work as dimension reduction methods.

3.1. PCA and EFM

Suppose a face image is represented by a feature vector $\mathbf{x}$, its total covariance matrix Σ_t and within-class covariance matrix Σ_w are defined in Eq. (1) and Eq. (2), respectively. $\mathbf{x}_{ij}$ denotes j-th sample of class i, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q_i$. p indicates the number of classes and q_i indicates the number of samples for class i. $\bar{\mathbf{x}}_i$ indicates the mean of training samples in class i and $\bar{\mathbf{x}}$ indicates the mean of all training samples and T indicates transpose.

$$\Sigma_t = \sum_{i=1}^{p} \sum_{j=1}^{q_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})^T. \tag{1}$$

$$\Sigma_w = \sum_{i=1}^{p} \sum_{j=1}^{q_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T. \tag{2}$$

PCA uses the Karhunen-Loeve Transform to produce the most expressive subspace for face representation and recognition. It factorizes Σ_t in Eq. (3) and obtain the eigenvector matrix Φ. Eigenvectors corresponding to d largest eigenvalues in Λ are used as the

projection matrix P in Eq. (4) to compute the d-dimensional vector $\mathbf{y}$ in the PCA subspace.

$$\Sigma_t = \Phi \Lambda \Phi^T. \tag{3}$$

$$\mathbf{y} = P^T \mathbf{x}. \tag{4}$$

In order to use the Mahalanobis distance for similarity comparison between $\mathbf{y}$ rather than the Euclidean distance, we compute the within-class covariance matrix Σ_{wy} of $\mathbf{y}$ according to Eq. (2). Eigenvector matrix Φ_{wy} and eigenvalue matrix Λ_{wy} of Σ_{wy} are derived similarly as in Eq. (3). Then the whitening matrix Q is computed in Eq. (5).

$$Q = \Phi_{wy} (\Lambda_{wy})^{-\frac{1}{2}} \tag{5}$$

The final d-dimensional vector $\mathbf{z}$ for distance comparison is

$$\mathbf{z} = Q^T P^T \mathbf{x} = U^T \mathbf{x}. \tag{6}$$

Enhanced Fisher Model [54] is an example of discriminating subspace methods. It achieves high separability among the different pattern classes. The first step of EFM is the same as PCA in Eq. (4). After that, EFM computes the within-class covariance matrix Σ_{wy}, and the between-class covariance matrix Σ_{by} of $\mathbf{y}$ which is computed according to Eq. (7).

$$\Sigma_{by} = \sum_{i=1}^{p} q_i (\overline{\mathbf{y}}_i - \overline{\mathbf{y}})(\overline{\mathbf{y}}_i - \overline{\mathbf{y}})^T. \tag{7}$$

Eigenvector matrix Φ_{wb} and eigenvalue matrix Λ_{wb} are derived by solving the eigenvalue problem below

$$\Sigma_{wy}^{-1} \Sigma_{by} = \Phi_{wb} \Lambda_{wb} \Phi_{wb}^T. \tag{8}$$

Then a projection matrix H consisting of eigenvectors in Φ_{wb} corresponding to d' largest eigenvalues in Λ_{wb} is used to compute the final d'-dimensional vector $\mathbf{z}$

$$\mathbf{z} = H^T P^T \mathbf{x} = U^T \mathbf{x}. \tag{9}$$

Many other dimension reduction methods are modifications or extensions of the above two methods. Thus PCA and EFM are taken as two representative dimension reduction methods used in this work.

3.2. DR-Cat approach

Let $\mathbf{x}^1, \ldots, \mathbf{x}^n, \ldots, \mathbf{x}^N$ be N vectors of different features extracted from the same face. DR-Cat approach computes the covariance matrix Σ^n for each feature vector $\mathbf{x}^n$ separately, where Σ^n can be a total covariance matrix, within-class covariance matrix or between-class covariance matrix. From Σ^n, the projection matrix U^n is derived to project the high-dimensional feature vector $\mathbf{x}^n$ to a low-dimensional feature vector $\mathbf{z}^n$ as in Eq. (6) of PCA or Eq. (9) of EFM. Note that the covariance matrix Σ^n provides only within-feature information, which means that the dimension reduction is implemented independently on each feature. Then low-dimensional feature vectors $\mathbf{z}^1, \ldots, \mathbf{z}^n, \ldots, \mathbf{z}^N$ are concatenated into $\mathbf{z}$ for classification as in equation (10). Each low-dimensional feature vector is normalized to have zero mean and unit variance prior to their concatenation.

$$\begin{aligned} \mathbf{z} &= [\mathbf{z}^1; \ldots; \mathbf{z}^n; \ldots; \mathbf{z}^N] \\ &= [(U^1)^T \mathbf{x}^1; \ldots; (U^n)^T \mathbf{x}^n; \ldots; (U^N)^T \mathbf{x}^N]. \end{aligned} \tag{10}$$

3.3. Cat-DR approach

Cat-DR approach concatenates different feature vectors $\mathbf{x}^n$ into an overall feature vector $\mathbf{x}$, $\mathbf{x} = [\mathbf{x}^1; \ldots; \mathbf{x}^n; \ldots; \mathbf{x}^N]$, to make use of their correlation information. Different feature vectors are normalized to have zero mean and unit variance before concatenation. Then the projection matrix U is derived from the covariance matrix Σ of the overall feature vector $\mathbf{x}$ as in Eq. (6) of PCA or Eq. (9) of EFM to project $\mathbf{x}$ to a low-dimensional feature vector $\mathbf{z}$ in equation (11) for classification.

$$\begin{aligned} \mathbf{z} &= U^T \mathbf{x} \\ &= U^T [\mathbf{x}^1; \ldots; \mathbf{x}^n; \ldots; \mathbf{x}^N]. \end{aligned} \tag{11}$$

4. Covariance matrix regularization for feature fusion

4.1. Covariance matrices in DR-Cat and Cat-DR

The projection matrices, U^n in DR-Cat and U in Cat-DR, are derived from the covariance matrices of training data, Σ^n in DR-Cat and Σ in Cat-DR, respectively. Σ^n are in fact submatrices of Σ. The covariance matrix carries two different kinds of information: data variances of the variables and the covariances between each pair of variables. Σ^n consists of data variances and covariances within feature $\mathbf{x}^n$ while Σ possesses both within-feature covariances and cross-feature covariances. For a better understanding, we represent the covariance matrix Σ as:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \ldots & \Sigma_{1n} & \ldots & \Sigma_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \ldots & \Sigma_{nn} & \ldots & \Sigma_{nN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{N1} & \ldots & \Sigma_{Nn} & \ldots & \Sigma_{NN} \end{pmatrix}. \tag{12}$$

As shown in Eq. (12), Σ can be represented as a block covariance matrix whose entries are partitioned into within-feature submatrices, denoted by $\Sigma_{nn}, n = 1, 2, \ldots, N$, which are the same as Σ^n in DR-Cat, and cross-feature submatrices, denoted by $\Sigma_{nm}, n \neq m, n, m = 1, 2, \ldots, N$, which are ignored in DR-Cat.

Within-feature submatrix Σ_{nn} is computed by feature vectors $\mathbf{x}^n$. It contains data variances and covariances within feature vector $\mathbf{x}^n$. Cross-feature submatrix Σ_{nm} contains data covariances between two different features $\mathbf{x}^n$ and $\mathbf{x}^m$. These cross-feature covariances have critical influence on the process of fusing different features. DR-Cat derives its projection matrix U^n from Σ_{nn}, ignoring the correlation between different features contained in Σ_{nm}. Cat-DR makes use of both Σ_{nn} and Σ_{nm} to derive U for feature fusion. In the ideal case of perfect training data which provides reliable and consistent information with the data population, Cat-DR achieves better performance than DR-Cat. However, in practice, the limited number of training samples may result in unreliable estimates of the cross-feature covariances, which causes overfitting and may make Cat-DR underperform DR-Cat.

4.2. Overfitting and covariance matrix regularization

Overfitting is a modelling error which occurs when a function is too closely fit to a limited set of training data. In reality, the data being studied often has some degree of noise or error within it. Thus making a model conform closely to inaccurate data can affect the model with substantial errors and reduce its predictive power. The degree of overfitting depends on the level of noise in the training data.

In general, Cat-DR should deliver better performance than DR-Cat as it takes account of the correlation information between different features. However, the correlation information is estimated from the training data, which usually deviates from that of the data population, especially in the case of limited number of training samples. When the feature fusion model is trained to closely conform to the estimated correlation information from the finite

training data, the resulting model will show overfitting and performance degradation on the data population or new data.

In order to reduce overfitting by regularization, authors in [29,30] add a constant to diagonal elements of the covariance matrix. Another solution is to decompose the discriminant function into two parts and replace the small eigenvalues of the covariance matrix by a constant as in [31,32]. Besides adding a constant to all eigenvalues or replacing the unreliable eigenvalues by a constant as discussed above, ERE in [33] replaces the unreliable eigenvalues with a model determined by the reliable eigenvalues. These three methods regularize the biased covariance matrix of training data by modifying its eigenvalues thus the regularized eigenspectrum can be closer to the population variances. However, modifying eigenvalues changes the trace of the covariance matrix and reduces the discriminating power of features themselves. In this paper, we propose a covariance matrix regularization (CMR) method to solve the problem of unreliable estimates of cross-feature correlations in feature fusion. Instead of modifying eigenvalues, it assigns weights w_{nm}, $0 < w_{nm} < 1$, to cross-feature submatrices Σ_{nm} in the covariance matrix Σ as shown below:

$$\Sigma^R = \begin{pmatrix} \Sigma_{11} & \dots & w_{1n} * \Sigma_{1n} & \dots & w_{1N} * \Sigma_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{n1} * \Sigma_{n1} & \dots & \Sigma_{nn} & \dots & w_{nN} * \Sigma_{nN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{N1} * \Sigma_{N1} & \dots & w_{Nn} * \Sigma_{Nn} & \dots & \Sigma_{NN} \end{pmatrix}. \tag{13}$$

The optimal value of w_{nm} depends on how much regularization is required for two different features $\mathbf{x}_n$ and $\mathbf{x}_m$, which can be estimated using some prior knowledge of feature properties and training data. For example, a relatively small weight is required in the case of large deviation between the estimated correlation and that of the data population. Experimental evaluation of the optimal value of weights in CMR can be found in Section 5.3. By using CMR, the influence of correlation information estimated from the training data is suppressed. The feature fusion model learns from but does not adapt too much to the estimated correlation thus increases its generalization ability to unknown instances.

When fusing features, the technique of CMR defined in Eq. (13) is applied to the total covariance matrix and the within-class covariance matrix of training data. It is straightforward to compute Σ_t^R from Σ_t defined in Eq. (1) according to Eq. (13). The within-class covariance matrix Σ_{wy} is computed in the PCA subspace, where different original features are mixed in the low-dimensional feature vectors $\mathbf{y}$. Therefore, the within-class covariance matrix in the PCA subspace can not be directly regularized as in Eq. (13). To solve this problem, we apply CMR to the within-class covariance matrix of original feature vectors $\mathbf{x}$, Σ_w defined in Eq. (2), to compute Σ_w^R according to Eq. (13). Then, we apply P^R, which consists of d largest eigenvectors of Σ_t^R, to Σ_w^R as in Eq. (14) to compute the regularized within-class covariance matrix in the PCA subspace Σ_{wy}^R.

$$\Sigma_{wy}^R = (P^R)^T \Sigma_w^R P^R. \tag{14}$$

Details of the proposed CMR method are summarized in Algorithm 1.

Algorithm 1 Covariance Matrix Regularization (CMR) using PCA or EFM.

1: Calculate the total covariance matrix Σ_t, and within-class covariance matrix Σ_w of $\mathbf{x}$ as in equation (1) and equation (2).
2: Apply CMR to Σ_t as in equation (13) and calculate P^R from Σ_t^R according to equation (3).
3: Apply CMR to Σ_w as in equation (13) and apply P^R to Σ_w^R to obtain Σ_{wy}^R as in equation (14).
4: Derive projection matrices using Σ_t^R and Σ_{wy}^R according to equation (6) for PCA or equation (9) for EFM.

5. Experiments

We assess the effectiveness of the proposed CMR technique for face recognition under 4 different feature fusion schemes: (1), fusion of pixel values from *R, G, B* channels; (2), fusion of LBP features from *R, G, B* channels; (3), fusion of pixel values and LBP features in the *R* channel; (4), fusion of CNN features extracted by VGG-Face and ResNetShort models. Extensive experiments are conducted on four publicly available face databases: MultiPIE [55], GT [56], AR [57], and LFW [58].

The Multi-PIE database contains face images captured under variations of illumination, poses and expressions in four recording sessions. We use the largest variation subset, illumination subset, which consists of 105 subjects with 80 face images per subject across 4 sessions (20 images per subject in each session). Similar to [59], we randomly choose s samples from 20 samples per subject in session 1 as the training and gallery data. Remaining 6300 face images of 105 subjects in session 2 to session 4 serve as query data. The nearest neighbor classifier with mahalanobis distance is used for classification. The gallery image is obtained by averaging all training samples per person. Face regions are cropped from original images and resized to the resolution of 32×32 for extraction of pixel values, LBP and CNN features. The patch size of LBP operator is set to be 4×4. Sample images are shown in Fig. 4.

The Georgia Tech (GT) [56] face database consists of 50 subjects with 15 images per subject. It characterizes several variations such as pose, expression, cluttered background, and illumination (see Fig. 5). Similar to Multi-PIE database, we randomly choose s samples from 15 samples per subject as the training and gallery data. Remaining $(15 - s)$ face images per subject serve as query data. The classifier is same as that used on Multi-PIE. The original images are downsampled to the size of 32×32 for extraction of pixel values, LBP and CNN features. The patch size of LBP is set to be 8×8.

The AR face database contains over 4000 color face images of 126 people, including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of most persons were taken in two sessions (separated by two weeks). In our experiments, 100 subjects with 14 frontal-face images per subject across 2 sessions (7 images per subject in each session) are selected. Only the full facial images were considered here (no attempt was made to handle occluded face recognition). In each session, there are 7 undisguised images with different facial expressions and lighting conditions for each subject. Similarly to before, we randomly choose s samples from 7 samples per subject in session 1 as the training and gallery data. Remaining 700 face images of 100 subjects in session 2 serve as query data. The classifier is same as that used on Multi-PIE. Face portions are manually cropped from original images and resized to the resolution of 32×32 for extraction of pixel values, LBP and CNN features. The patch size of LBP operator is set to be 8×8. Sample images are shown in Fig. 6.

The LFW database contains 13,233 images of 5749 subjects. Images in this database exhibit rich intra-personal variations of pose, illumination, and expression. It has been extensively studied for the research of unconstrained face recognition in recent years. In this paper, we follow the "Unrestricted, Labeled Outside Data Results" protocol defined in [58] and report the mean verification accuracy by the 10-fold cross-validation scheme on the View 2 data. Face images are normalized and aligned using the same method on

Fig. 4. Sample face images from the illumination variation subset of Multi-PIE database.

Fig. 5. Sample face images from Georgia Tech face database.

Fig. 6. Sample face images from AR database.

Fig. 7. Sample face images from LFW database.

CASIA-WebFace images. For the face verification paradigm, we use the CNN features provided by the VGG-Face model and the Resnet-Short model as the raw representation of each test sample. Then, we adopt cosine distance for similarity calculation (Fig. 7).

A number of experiments are conducted. To begin with, we evaluate the face verification performance of the pre-trained VGG-Face model and our proposed ResNetShort model on the challenging LFW dataset. Then, by decreasing the value of weights from 1 to 0, we validate that CMR solves the overfitting problem and improves the face recognition performance. After that, we show that with the number of training samples per subject decreasing, stronger regularization should be applied to the estimated cross-feature covariances by using smaller weights in CMR. Finally, the face recognition performance of the proposed CMR technique is compared with that of the best single feature, DR-Cat and Cat-DR by fusing features from multiple color channels, multiple types of features, and features extracted by different deep models. For the convenience and clarity of all experiments, we adopt the same value w for w_{nm}, $w_{nm} = w$, in CMR.

5.1. Performance evaluation of ResNetShort features

In this experiment, the face verification performance of ResNet-Short is evaluated on the challenging LFW database. Following the "Unrestricted, Labeled Outside Data Results" protocol, we input aligned face images to ResNetShort models and take the output of the first fully-connected layer as the deep features. The unsupervised diagram is used here, where PCA and cosine distance are used to calculate the similarity between two CNN features. We evaluate the covariance matrix of CNN features for PCA using the 9 training folds of LFW data in the 10-fold cross validation. The face verification performances of VGG-Face and ResNetShort are reported in Table 2, we also compare them with other state-of-the-art models DeepID [60] and Canonical View CNN [61], which have been peer-reviewed and published. FaceNet [41] is not included in Table 2 for comparison, as it is trained from 260M images of 8M subjects and uses a complex triplets selection algorithm. It is not fair to compare it with other deep models trained using less than 0.5M images. We can observe from Table 2, the proposed ResNet-Short model achieves comparable performance to other CNN models.

5.2. Evaluation of CMR against the level of regularization

Here, we conduct experiments to investigate how different levels of covariance matrix regularization make influence on the face recognition performance. Specifically, we vary the value of weights w in CMR from 1 to 0, so that its regularizing effect on the cross-feature covariances changes from weaker to stronger.

This experiment is carried out on Multi-PIE, GT and AR datasets. Face images in different color channels are arranged into column vectors as features of pixel values and LBP features are extracted from different channels separately. The radius and the number of sampling points in the LBP operator are set as 1 and 8 through our paper. For features of pixel values and LBP, the numbers of training samples per subject s are 4 on Multi-PIE, 5 on GT and 4 on AR. For CNN features, s equals 2 on Multi-PIE, GT and AR to increase the difficulty of the face recognition task.

We report the face recognition performances of different weights in CMR by face recognition rate (FRR), which is the ratio of the number of correctly classified query images to the total number of query images. Note that, among all tested feature dimensions of PCA and EFM, the best found FRR is reported. We plot FRR against the value of weights in CMR for 4 different feature fusion schemes on Fig. 8 and Fig. 9. As we can observe, when the value of weights in CMR decreases from 1 to zero, the FRR

Table 2
Face verification accuracy (%) of ResNetShort, VGG-Face, DeepID and Canonical CNN on LFW.

Model	ResNetShort	VGG-Face	DeepID [60]	Can. CNN [61]
Verif. metric	Cosine	Cosine	Joint Bayes	Joint Bayes
Mean accuracy (%)	98.72	97.93	97.45	96.45

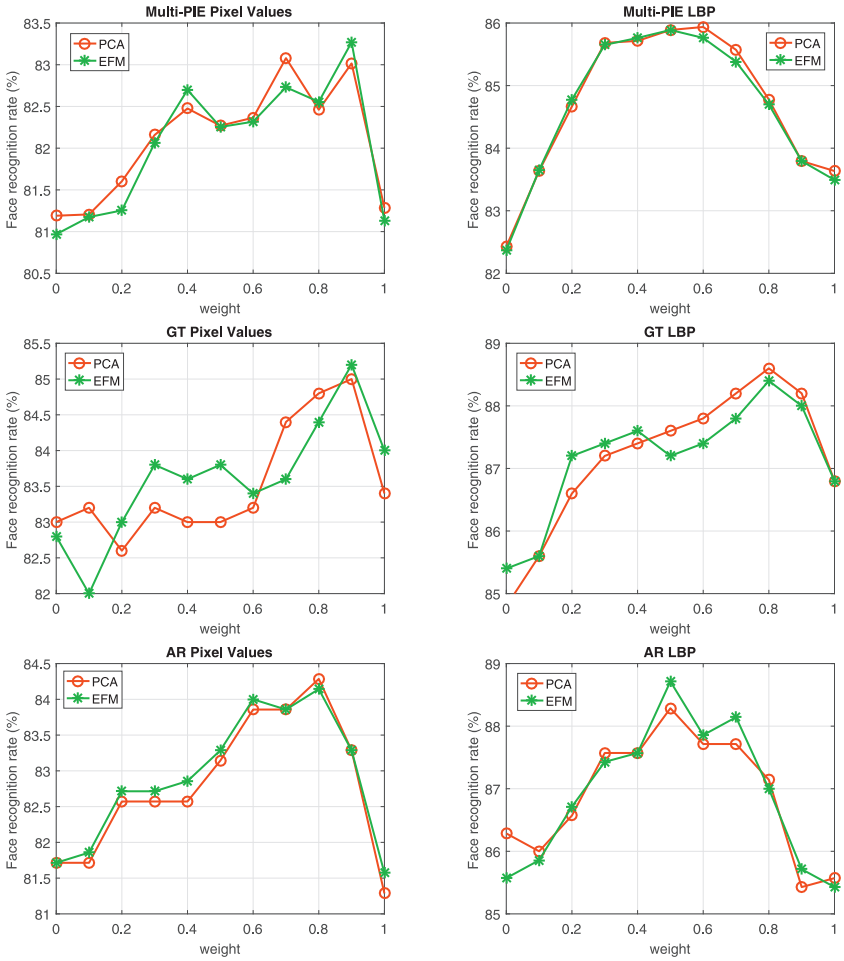

Fig. 8. Face recognition rates (%) of fusing features (pixel values or LBP) of 3 color channels (R,G,B) against the value of weights in CMR on Multi-PIE, GT and AR. Each column specifies one type of feature (pixel values or LBP) and each row specifies one dataset (Multi-PIE, GT and AR).

increases to the maximum point and then decreases for all 4 feature fusion schemes on all three databases. The best performance is achieved at weight of 0.5 ~ 0.9. One clear and consistent conclusion summarized from Fig. 8 and Fig. 9 is that, applying CMR to feature fusion improves the face recognition performance consistently on all 4 feature fusion schemes and all 3 datasets.

5.3. The optimal value of weights in CMR for different training data

In this section, we investigate how the optimal value of weights in CMR will change with the decreasing of the number of training samples per subject. The experiment is conducted on GT and AR datasets, where CMR is used for fusing LBP features of different color channels (R, G, B) and PCA is used for dimension reduction. The FRR of CMR trained by s samples per subject are plotted against the regularization parameter w on Fig. 10 for GT and on Fig. 11 for AR. On GT, s=10, 5, 3 are tested. On AR, s=6, 5, 4 are tested. As we can observe from Fig. 10 and Fig. 11, when the number of training samples per subject decreases, the optimal value of weights in CMR (indicated by dotted lines) that achieves the best face recognition performance also deceases. As fewer training samples per subject are provided to the feature fusion model, the estimated cross-feature covariances from training data are less reliable and hence need more regularization. Thus lower weights should be assigned to the cross-feature covariances in CMR with smaller size of training data provided.

5.4. Performance comparison of CMR against the best single feature, DR-Cat and Cat-DR

To systematically compare the performance of CMR with that of the best single feature, DR-Cat and Cat-DR, we conduct experiments on Multi-PIE, GT, AR and LFW datasets. In CMR, we vary the value of w from 0 to 0.9 with step size of 0.1 and report the best classification performance. Training and testing protocols of Multi-PIE, GT and AR are the same as those in Section 5.2. To increase the difficulty of the face verification task on LFW, only 1 training fold in the 10-fold cross validation is used to train PCA or EFM, remaining 9 training folds are used for testing. Other experimental settings remain the same as in Section 5.1. We show

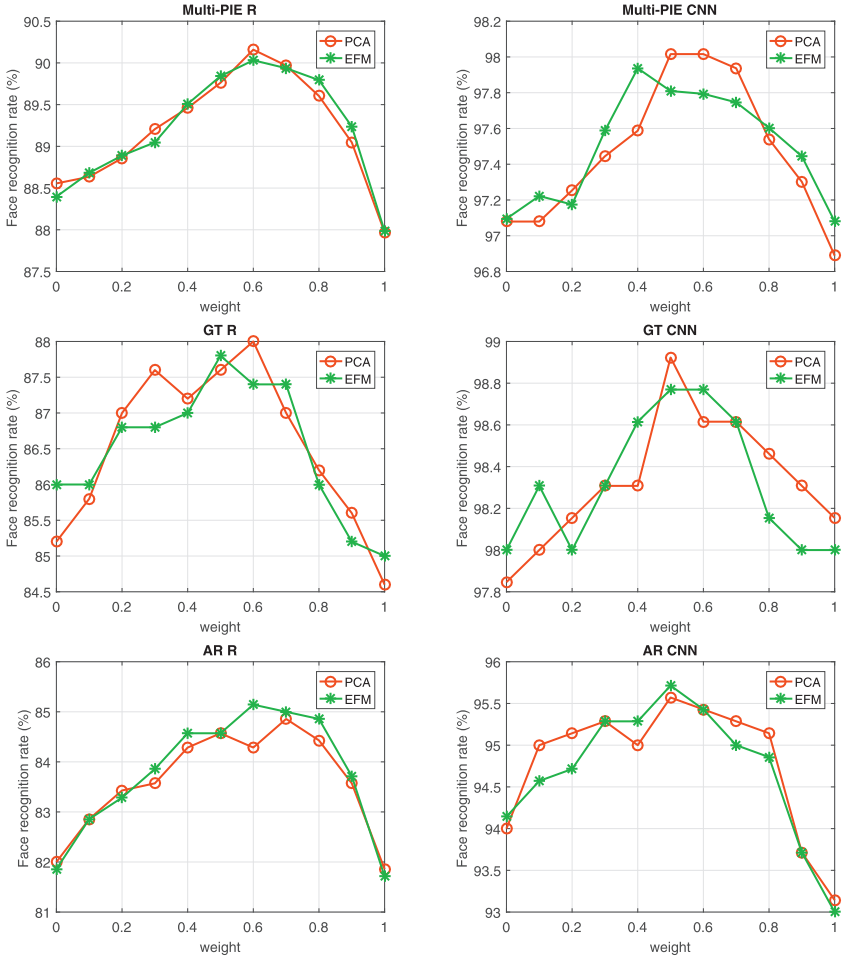

Fig. 9. Face recognition rates (%) of fusing different types of features (pixel values and LBP of channel *R*) and fusing features extracted by different deep models (VGG-Face and ResNetShort) against the value of weights in CMR on Multi-PIE, GT and AR. Each column specifies one type of feature fusion (multi-type or multi-model) and each row specifies one dataset (Multi-PIE, GT and AR).

Table 3
Face recognition performances of the best single feature, DR-Cat, Cat-DR and CMR using pixel values of multiple color channels on Multi-PIE, GT and AR.

Pixel values	Multi-PIE		GT		AR	
	PCA	EFM	PCA	EFM	PCA	EFM
B.S. channel	81.73	81.95	79.20	79.60	79.86	80.00
DR-Cat	80.95	80.73	81.00	81.40	83.14	82.86
Cat-DR	81.29	81.13	83.40	84.00	81.29	81.57
CMR	**83.08**	**83.27**	**85.00**	**85.20**	**84.29**	**84.14**

Table 4
Face recognition performances of the best single feature, DR-Cat, Cat-DR and CMR using LBP of multiple color channels on Multi-PIE, GT and AR.

LBP	Multi-PIE		GT		AR	
	PCA	EFM	PCA	EFM	PCA	EFM
B.S. channel	82.40	82.24	79.20	79.60	80.71	81.14
DR-Cat	84.79	84.71	86.60	86.80	85.86	86.14
Cat-DR	83.63	83.49	86.80	86.80	85.57	85.43
CMR	**85.94**	**85.89**	**88.20**	**88.00**	**88.29**	**88.71**

Table 5
Face recognition performances of the best single feature, DR-Cat, Cat-DR and CMR using pixel values and LBP of channel *R* on Multi-PIE, GT and AR.

R	Multi-PIE		GT		AR	
	PCA	EFM	PCA	EFM	PCA	EFM
B.S. feature	83.08	82.83	83.60	83.40	78.57	78.86
DR-Cat	88.51	88.29	86.20	86.60	82.86	83.14
Cat-DR	87.97	87.98	84.60	85.00	81.86	81.71
CMR	**90.16**	**90.03**	**88.00**	**87.80**	**84.86**	**85.14**

the FRR of the best single (B.S.) feature, DR-Cat, Cat-DR and CMR on Tables 3–6 for the fusion of pixel values of *R*, *G*, *B* channels, the fusion of LBP of *R*, *G*, *B* channels, the fusion of pixel values and LBP of channel *R*, and the fusion of CNN features extracted by VGG-Face and ResNetShort, respectively. We use **bold** texts and underline texts to highlight the highest and the second highest face recognition/verification accuracy among all methods, respectively.

As shown in Tables 3–6, the best single feature performs worse than all feature fusion methods (DR-Cat, Cat-DR and CMR) in 22 of the 26 experiments, which indicates that the fusion of multiple features is effective in promoting the face recognition performance. Although Cat-DR should perform better than DR-Cat in the ideal situation, it outperforms DR-Cat only in 11 of the 26 experiments. This shows that the full use of correlation information estimated from the training data causes overfitting that reduces the predictive accuracy. We propose the CMR technique to solve problems in DR-Cat and Cat-DR. In CMR, the correlation information is regularized and then used to train the feature fusion model. Results show that the proposed CMR consistently performs better than the best

Table 6
Face recognition/verification performances of the best single model, DR-Cat, Cat-DR and CMR using CNN features of multiple deep models on Multi-PIE, GT, AR and LFW.

CNN	Multi-PIE		GT		AR		LFW	
	PCA	EFM	PCA	EFM	PCA	EFM	PCA	EFM
B.S. model	94.89	95.08	97.85	97.38	91.43	91.57	97.14	97.14
DR-Cat	96.89	96.65	94.92	94.62	93.86	94.43	97.56	97.42
Cat-DR	96.89	97.08	98.15	98.00	93.14	93.00	97.28	97.44
CMR	**98.02**	**97.94**	**98.92**	**98.77**	**95.57**	**95.71**	**97.80**	**97.83**

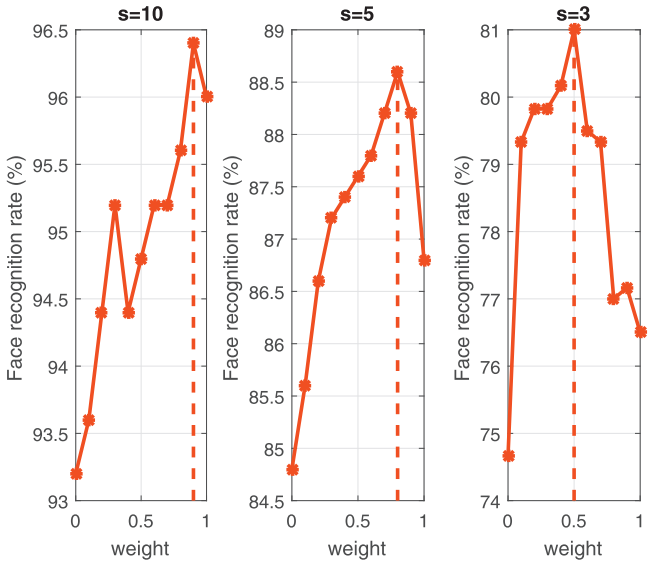

Fig. 10. Face recognition rates against the value of weights of CMR for different numbers (s) of samples per subject on GT.

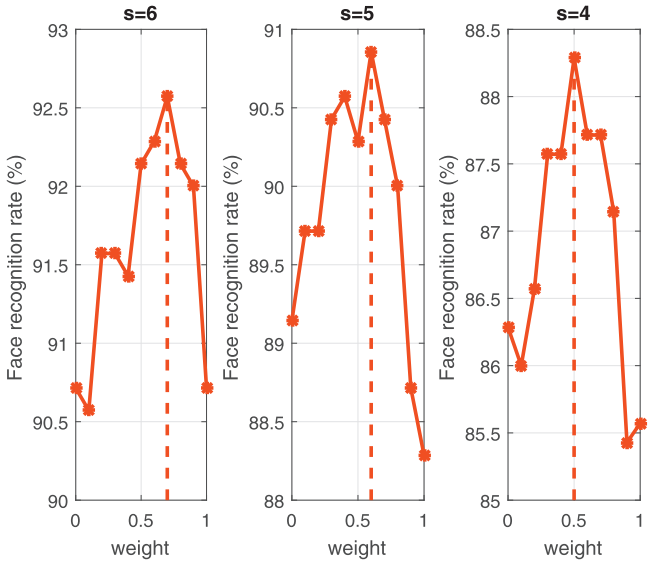

Fig. 11. Face recognition rates against the value of weights of CMR for different numbers (s) of samples per subject on AR.

single feature, DR-Cat and Cat-DR for fusing features of different color channels, different types of features and features extracted by different deep models in all 26 experiments.

6. Conclusion

In this paper, we propose a covariance matrix regularization (CMR) technique to utilize the correlation between different features and reduce overfitting during the fusion of multiple features. It works by assigning weights to the cross-feature submatrices of covariance matrices of training data to suppress the influence of correlation between different features, which is estimated from the training data, in feature fusion. Extensive experiments conducted on four popular face databases show that the proposed CMR technique consistently outperforms the best single feature, DR-Cat and Cat-DR for fusing features of different color channels, different types of features and features extracted by different deep models.

Acknowledgments

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. This research was partly supported by Singapore Ministry of Education Academic Research Fund Tier 1 RG 123/15. The ROSE Lab is supported by the National Research Foundation, Singapore, under its Interactive Digital Media (IDM) Strategic Research Programme.

References

[1] Y.K. Park, S.L. Park, J.K. Kim, Retinex method based on adaptive smoothing for illumination invariant face recognition, Signal Process. 88 (8) (2008) 1929–1945.
[2] Z. Zheng, F. Yang, W. Tan, J. Jia, J. Yang, Gabor feature-based face recognition using supervised locality preserving projection, Signal Process. 87 (10) (2007) 2473–2483.
[3] T. Mandal, Q.J. Wu, Y. Yuan, Curvelet based face recognition via dimension reduction, Signal Process. 89 (12) (2009) 2345–2353.
[4] B. Xiao, X. Gao, D. Tao, X. Li, A new approach for face recognition by sketches in photos, Signal Process. 89 (8) (2009) 1576–1588.
[5] D.V. Jadhav, R.S. Holambe, Radon and discrete cosine transforms based feature extraction and dimensionality reduction approach for face recognition, Signal Process. 88 (10) (2008) 2604–2609.
[6] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition, Comput. Vis. Image Understanding 101 (1) (2006) 1–15.
[7] C. Liu, H. Wechsler, A shape-and texture-based enhanced fisher classifier for face recognition, IEEE Trans. Image Process. 10 (4) (2001) 598–608.
[8] X. Tan, B. Triggs, Fusing Gabor and lbp feature sets for kernel-based face recognition, in: International Workshop on Analysis and Modeling of Faces and Gestures, Springer, 2007, pp. 235–249.
[9] Z. Liu, C. Liu, Fusion of the complementary discrete cosine features in the YIQ color space for face recognition, Comput. Vis. Image Understanding 111 (3) (2008) 249–262.
[10] Z. Liu, C. Liu, Fusion of color, local spatial and global frequency information for face recognition, Pattern Recognition 43 (8) (2010) 2882–2890.
[11] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Color local texture features for color face recognition, IEEE Trans. Image Process. 21 (3) (2012) 1366–1380.
[12] S.H. Lee, J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Local color vector binary patterns from multichannel face images for face recognition, IEEE Trans. Image Process. 21 (4) (2012) 2347–2353.
[13] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Trans. Image Process. 19 (6) (2010) 1635–1650.
[14] Y. Su, S. Shan, X. Chen, W. Gao, Hierarchical ensemble of global and local classifiers for face recognition, IEEE Trans. Image Process. 18 (8) (2009) 1885–1896.
[15] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: IEEE Conf. Computer Vision and Pattern Recognition, 2014.
[16] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, IEEE Trans. Multimedia 17 (11) (2015) 2049–2058.

[17] Z. Tu, J. Cao, Y. Li, B. Li, MSR-CNN: applying motion salient region based descriptors for action recognition, in: Intl. Conf. Pattern Recognition, IEEE, 2016, pp. 3524–3529.
[18] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, 1, 2015, p. 6.
[19] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[20] F. Gurpinar, H. Kaya, H. Dibeklioglu, A. Salah, Kernel ELM and CNN based facial age estimation, in: The IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops, 2016.
[21] K. Zhang, L. Tan, Z. Li, Y. Qiao, Gender and smile classification using deep convolutional neural networks, in: The IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops, 2016.
[22] X. Peng, Z. Xia, L. Li, X. Feng, Towards facial expression recognition in the wild: a new database and deep recognition system, in: IEEE Conf. Computer Vision and Pattern Recognition Workshops, 2016, pp. 93–99.
[23] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, arXiv preprintarXiv:1411.7923(2014).
[24] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification, in: IEEE Conf. Computer Vision and Pattern Recognition, 2013, pp. 3025–3032.
[25] Z. Lu, X. Jiang, A.C. Kot, A color channel fusion approach for face recognition, IEEE Signal Process. Lett. 22 (11) (2015) 1839–1843.
[26] Z. Liu, C. Liu, A hybrid color and frequency features method for face recognition, IEEE Trans. Image Process. 17 (10) (2008) 1975–1980.
[27] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Boosting color feature selection for color face recognition, IEEE Trans. Image Process. 20 (5) (2011) 1425–1434.
[28] J. Yang, C. Liu, Color image discriminant models and algorithms for face recognition, IEEE Trans. Neural Netw. 19 (12) (2008) 2088–2098.
[29] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, Pattern Recognit. Lett. 26 (2) (2005) 181–191.
[30] S. Ji, J. Ye, Generalized linear discriminant analysis: a unified framework and efficient model selection, Trans. Neural Netw. 19 (10) (2008) 1768–1782.
[31] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 696–710.
[32] B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition, Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 780–788.
[33] X. Jiang, B. Mandal, A. Kot, Eigenfeature regularization and extraction in face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 30 (3) (2008) 383–394.
[34] P. Shih, C. Liu, Comparative assessment of content-based face image retrieval in different color spaces, Int. J. Pattern Recognit. Artif. Intell. 19 (07) (2005) 873–893.
[35] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Color face recognition for degraded face images, IEEE Trans. Systems Man Cybern. B: Cybern. 39 (5) (2009) 1217–1230.
[36] J. Yang, C. Liu, L. Zhang, Color space normalization: enhancing the discriminating power of color spaces for face recognition, Pattern Recognit. 43 (4) (2010) 1454–1466.
[37] C. Liu, Learning the uncorrelated, independent, and discriminating color spaces for face recognition, IEEE Trans. Inf. Forensics Secur. 3 (2) (2008) 213–222.
[38] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[39] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, 2015.
[40] M. Mathias, R. Benenson, M. Pedersoli, L. Van G., Face detection without bells and whistles, in: European Conf. Computer Vision, Springer, 2014, pp. 720–735.
[41] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: IEEE Conf. Computer Vision and Pattern Recognition, 2015, pp. 815–823.
[42] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: European Conf. Computer Vision, Springer, 2016, pp. 499–515.
[43] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: IEEE International Conf. Computer Vision, 2015, pp. 1026–1034.
[44] E. Zhou, Z. Cao, Q. Yin, Naive-deep face recognition: touching the limit of LFW benchmark or not?, arXiv preprintarXiv:1501.04690(2015).
[45] Z. Zhang, P. Luo, C.C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, in: European Conference on Computer Vision, Springer, 2014, pp. 94–108.
[46] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: ACM International Conf. Multimedia, 2014, pp. 675–678.
[47] R.O. Duda, P.E. Hart, D.G. Stork, et al., Pattern classification, 2, Wiley New York, 1973.
[48] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.
[49] D. Pissarenko, Eigenface-based facial recognition, December 1st (2002).
[50] P.J. Phillips, H. Moon, S. Rizvi, P.J. Rauss, et al., The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.
[51] X. Jiang, Linear subspace learning-based dimensionality reduction, IEEE Signal Process. Mag. 28 (2) (2011) 16–26.
[52] X. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 931–937.
[53] J. Yang, C. Liu, A discriminant color space method for face representation and verification on a large-scale database, in: IEEE Intl. Conf. Pattern Recognition, 2008, pp. 1–4.
[54] C. Liu, H. Wechsler, Robust coding schemes for indexing and retrieval from large face databases, IEEE Trans. Image Process. 9 (1) (2000) 132–137.
[55] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.
[56] A. Nefian, Georgia Tech face database, 2013.
[57] A.M. Martinez, The AR face database, CVC Tech. Rep. 24 (1998).
[58] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report, Technical Report 07–49, University of Massachusetts, Amherst, 2007.
[59] P. Zhu, L. Zhang, Q. Hu, S.C. Shiu, Multi-scale patch based collaborative representation for face recognition with margin distribution optimization, in: European Conference on Computer Vision, Springer, 2012, pp. 822–835.
[60] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Advances in Neural Information Processing Systems, 2014, pp. 1988–1996.
[61] X. Zhu, Z. Lei, J. Yan, D. Yi, S.Z. Li, High-fidelity pose and expression normalization for face recognition in the wild, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015.