



Face recognition in a large dataset using a hierarchical classifier

Navid Abbaspoor¹ · Hamid Hassanpour¹ 

Received: 28 September 2020 / Revised: 15 January 2022 / Accepted: 21 January 2022 /
Published online: 3 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Face recognition is one of the most common authentication methods. Although much research has been conducted in this area, there are still many challenging issues to be addressed on face recognition, such as a large number of images in a dataset, with only one sample per person. The goal of this paper is to provide a robust face recognition method for a database having a large number of images with only one sample per person. The proposed method first uses a simple clustering approach to divide the images hierarchically into balanced clusters. Balanced clustering helps us to continue clustering in several hierarchies and finally reach very small clusters of equal size. Then, the face recognition task is performed within each cluster. A combination of the Non-negative Matrix Factorization (NMF) and the Fast Retina Key-point (FREAK) descriptors has been used to match the faces. The proposed method was evaluated on the FERET dataset that achieved an accuracy of 98.36%. Also, some other experiments have been done to validate the efficiency of the proposed method. The results of the experiments show that the proposed method can be applied to even larger datasets, while its complexity increases linearly.

Keywords Face recognition · Large datasets · FERET dataset · Non-negative matrix factorization · Shi-Tomasi · FREAK

1 Introduction

The use of biometric authentication methods has been growing in recent decades. These methods use biometric factors such as iris, fingerprint, voice, and face for authentication [58]. However, the face is known as one of the popular biometric factors. The reason is the simplicity of using the face image and also the fact that there is no need for interaction between the person and the system. Face recognition is frequently used in video surveillance

✉ Hamid Hassanpour
h.hassanpour@shahroodut.ac.ir

Navid Abbaspoor
navid.abbaspoor@gmail.com

¹ Shahrood University of Technology, Shahrood, Semnan, Iran

[2], law enforcement [11], access control systems, facial rejuvenation [7] and so forth due to its ease of use as well as the possibility of remote authentication [29].

Despite the many advances that have been made in face recognition, it encounters many challenges, that each method is trying to solve one or more of them. Conditions such as illumination and changes in the pose of the face, occlusion and expressions [20] make it difficult to build a robust face recognition system [4]. Although some methods have worked these challenges, they require high processing time, high memory consumption, and are relatively complex [29, 32].

In addition to the above mentioned challenges, other factors such as large number of images in the dataset [44] and having only one sample per person [35] are also problematic. Most face recognition methods are not scalable and their performance dramatically decreases as the number of images increases. Also, most methods require several training samples for identification and fail if the number of samples is small [10]. The goal of this paper is to present a novel approach addressing problems of large number of images in the dataset and having only one sample per person.

The process of recognizing faces consists of three main steps: *Preprocessing*, *Feature extracting*, *Matching features* [32]. The first step is to divide the image into two segments, face and non-face [68], and also after detecting face, the image quality is enhanced in terms of brightness, contrast, etc. In the second step, discriminative features are extracted from face images to make it easier to identify faces. Finally, in the last step, extracted features are learned by classifiers to predict the identity of the new faces.

Face recognition methods are mainly divided into three categories based on their perspective: *Holistic*, *Local*, and *Hybrid* [29]. The first category, also known as subspace methods (such as Principal Components Analysis (PCA) [61], Linear Discriminant Analysis (LDA) [53], Independent Components Analysis (ICA) [6]), projects the entire image to a smaller space. The second category can be divided into two subcategories, local-appearance-based methods (such as Local Binary Patterns (LBP) [1] and its variations, Histogram of Oriented Gradients (HOG) [18]) that divide the images into blocks and extract features locally in each block, and the key-point-based methods (such as Scale-Invariant Feature Transform (SIFT) [15], Speeded-Up Robust Features (SURF) [75]) that detect points of interest in images, then extract local features around those points. The third category uses a combination of holistic and local methods (such as Gabor-LBP-LDA [8], DCT-PCA-GA [41]) to obtain better performance in recognition [60].

The purpose of this paper is to focus on the challenge of large dataset size and the Single Sample Per Person (SSPP) problem. Since the problem of face recognition can be considered as a multiple optimization problem [32], the existing methods usually focus on a single problem to solve. For example, Deep Convolutional Neural Network (DCNN) can work well with large datasets [51, 56, 66]. But they also have certain problems and limitations. Deep learning methods require a large number of training samples to work well [30]. They have high resource consumption and need too much time to be trained. These methods have many hyperparameters that must be specified [27] and also they are always at risk of overfitting [65].

Some other methods use clustering techniques to divide images into clusters in order to solve the problem of large dataset size. In this case, face recognition can be done with a relatively small number of images within each cluster [44, 46]. These methods are also associated with problems such as finding the right clustering method with high performance, and sometimes the need for a large number of samples for training.

To solve the SSPP problem, some methods project the images into another space where the distance between the samples is maximum [42, 71]. Another set of methods use an auxiliary dataset to learn the original gallery images with help of the generic set images [25, 26, 69]. Some other methods divide images into several patches and consider the extracted features of each patch as a new sample for each person [10, 37]. Other methods transform images to create new images in different spaces of the sample image to generate new samples for each person and use them for identification [38, 72].

This paper proposes a hierarchical face recognition method that is robust to large datasets and single sample per person. The proposed method uses a simple clustering approach to hierarchically divide images into clusters with fewer images. The clustering method uses a user-defined threshold to cluster images. Within each cluster, two sets of features are extracted using Non-negative Matrix Factorization (NMF) [28] and Fast Retina Keypoint (FREAK) [3] descriptors. NMF is one of the holistic methods that projects the entire image space to a feature space with smaller dimensions. Then it extracts basic features that are very similar to the human visual system [31]. FREAK descriptor is one of the local key-point-based methods that extracts binary features from the region of interests. Finally, these two set of features will be used to identify new faces using the nearest neighbor with City-Block distance measure. The proposed method combines these two features to take advantage of both holistic and local methods. The results of experiments performed on the FERET [47] dataset and an extended image set collected from other standard datasets indicate the good performance of the proposed method.

The main contributions of this paper include 1) a robust face recognition approach that solves the challenge of having a high number of images in the dataset. Using a simple clustering method, the images are divided into smaller sets, so the identification task can be done on smaller image sets. 2) Using one sample per person. Combining the Non-negative matrix factorization method and FREAK descriptors helps us to extract discriminative features with only one sample per person. The proposed method investigates on these two problems simultaneously. While previous works could only solve one of the problems at a time. 3) The proposed method does not require high-resolution images. The discriminative features can be extracted using 120×120 pixel images.

The rest of this paper is organized as follows. Section 2 summarizes some of the methods related to the subject of this paper. The proposed method is described in details in Section 3. To evaluate the proposed method, experiments have been performed, the explanations and results of these experiments are presented in Section 4. We then conclude the proposed method and discuss about the potential future directions in Section 5.

2 Related works

This section examines several face recognition methods that have been proposed to address the challenges of large dataset size or SSPP. Also, some other related works that use hierarchical approaches and hybrid features for image or face applications are examined. Lu et al. [36] proposed a method based on the LDA approach. In this method, to deal with face recognition on large-scale datasets first a clustering method is used to divide the images into smaller size sets. This method uses a “separability criterion” to partition the training set into a set of K smaller and simpler subsets or Maximal-Separability Clusters (MSCs). Then, inside the clusters, feature extraction is done using the LDA method. In the method

presented by Dugelay et al. [46] like Lu's method [36], images are clustered before classification. The difference is that the Maximum Likelihood (ML) and Expected-Maximization (EM) is used for clustering.

Choi's method [12] is one of the methods based on hierarchical classification. This method uses a combination of the wrinkle features extracted using Gabor filters and the skin features extracted using a local binary pattern. A hierarchical classifier based on a Support Vector Machine (SVM) and a Support Vector Regression (SVR) is used in this method. Some other methods used Radial Basis Function (RBF) networks to build a hybrid and hierarchical classifier [9, 17]. An ensemble of the RBF networks is used to build a hierarchical classifier. These methods were evaluated on a low number of subjects.

FaceNet [51] is a deep learning model that directly learns to map images to a compact Euclidean space. In this space, the distance between the images is directly related to their similarity. This model has been trained using about 260 million images in 1000 hours. This method has been evaluated on the Labeled Faces in the Wild (LFW) dataset for face verification tasks, which has good results.

Otto et al. [44] proposed a method based on Rank-Order clustering to cluster millions of faces by identity. LFW dataset images were used along with 123 million other unlabeled images. The Rank-Order method is a form of agglomerative hierarchical clustering that uses the nearest neighbor approach. This method starts with each image in its own cluster and starts to merge the two closest clusters. The minimum distance between any members in the clusters is considered as the distance between two clusters.

Zhang et al. [72] proposed two methods based on Singular Value Decomposition (SVD) perturbation to deal with the SSPP problem. In the first method, the original image is linearly combined with its derived image gotten by perturbing the image matrix's singular values. In the second method, the derived images are regarded as independent images that could augment the training image set. After all, PCA is performed on the joined images and derived images.

Lu et al. [37] proposed a Discriminative Multi-Manifold Analysis (DMMA) method that extracts discriminant features from image patches. To do this, the image is first divided into several non-overlapping patches to generate a set of images for each person. This image set is then used to extract the features and face recognition tasks. Liu et al. [35] used a modified K-means method to cluster two groups of people. Then they used the DMMA method to extract features from images. These two steps were repeated to obtain a binary tree projection matrix of fast DMMA. Finally, this tree is used for identification.

Xin et al. [67] created new images using geometric transformation from original images. Then the main images along with the new images are considered as a set of training images and feature extraction is done using the Local Gabor Binary Patterns (LGBP) method. Finally, the Locality Preserving Discriminant Analysis (LPDA) method is used to reduce the dimension of the features. Then the identity of the images is determined using the nearest neighbor method with Euclidean distance.

Hu et al. [24] proposed a method that uses lower-upper (LU) decomposition to generate two new image sets of original images. Then two approximation images were reconstructed from the two basis image sets. The authors used the Fisher Linear Discriminant Analysis (FLDA) method to extract features from the original image and its two approximation images.

Pan et al. [45] proposed an algorithm that is based on Locality Preserving Projection (LPP) to solve the SSPP problem. In this algorithm, first, transfer sources are screened to

obtain the selective sample source using the whitened cosine similarity metric. Then vectors of source faces and target faces are projected into feature subspace by LPP. Finally, the feature transfer matrix can be calculated to approximate the mapping relationship on the source and target faces.

Based on the transfer learning method of Zeng et al. [71], new samples are created using a traditional sample expansion approach. New samples are then used to fine-tune the DCNN model. The Class-Level Joint Representation with Regional Adaptive Convolution Features (CJR-RACF) method [70] was proposed by Yang et al. to solve the SSPP problem. This method learns the local and global features of images that are robust to facial variations. Min et al. [40] also used transfer learning to train a deep model using a large-scale dataset. Then, a sample expansion method in the feature space called K-Class Feature Transfer (KCFT) was introduced to enrich intra-class variation information for each image. Finally, expanded samples are used to train the classifier.

Nikan and Hassanpour [42] proposed a solution to the problem of large-scale datasets and SSPP based on NMF. In this method, images are first divided into two parts, top and bottom, then the basic features are extracted separately from the lower and upper parts, using the NMF method. Finally, the same images from the same identities are identified using the correlation distance measure.

In recent years, several deep learning methods for face recognition have been introduced. For example, DeepFace [57], DeepID [55], and ResNet [23] models, which were among the first deep methods used to identify and verify faces in unconstrained environments, performed well. However, these methods were very complex [22]. The next generation of deeper models such as Face++ [76] and FaceNet [51], which also had deeper structures, were able to improve the performance of previous methods as well.

The state-of-the-art in deep learning is the ArcFace [14] method, which uses the ResNet [23] structure and the Additive Angular Margin loss function to learn the face images. These approaches have even more accuracy than humans, which can identify a face among millions of images [51]. However, the training process requires a large number of samples from each person [30]. The resolution of the images strongly affects the performance of these models. These methods have complex computations and need a long time to be trained [27].

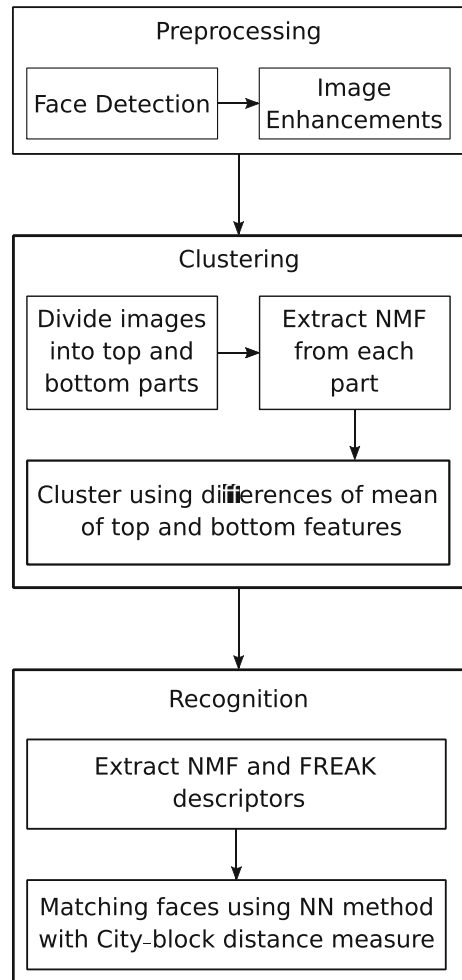
3 Proposed method

In general, the proposed method consists of three main steps: *preprocessing*, *clustering*, and *face recognition*. The face images are extracted from the original images. They are improved in terms of brightness and contrast. Then a simple method is proposed for clustering images into two clusters using NMF features. Finally, in each cluster, two types of features including NMF and FREAK descriptors are extracted and matching faces is done using the nearest neighbor method [13]. The whole process of the proposed method is drawn in Fig. 1.

3.1 Preprocessing

Since images may contain non-face regions, it is necessary to specify the location of the face in the image in the first step. Face detection at this stage is done using the known Viola-Jones [62] algorithm. We also examine another state-of-the-art face detection method

Fig. 1 Diagram of the proposed method



known as Multi-Task Cascaded Convolutional Networks (MTCNN) [73] in our experiments that results show its weaker performance than the traditional Viola-Jones method.

Then, the brightness and contrast of the extracted faces are improved by using conventional enhancement methods including Histogram Equalization (HE) [54], and Contrast Limited Adaptive Histogram Equalization (CLAHE) [49].

3.2 Clustering

Extracted face images after enhancements are ready for clustering. For this goal, we propose an innovative and simple approach. The images are first divided into two equal non-overlapping blocks, upper and lower parts. The upper part includes the eyes, eyebrows, and forehead, and the lower part includes the nose, lips, and chin. These steps are shown in Fig. 2. In order to achieve a better clustering, discriminative features should be extracted from images. Then, the features will be extracted separately from both parts.

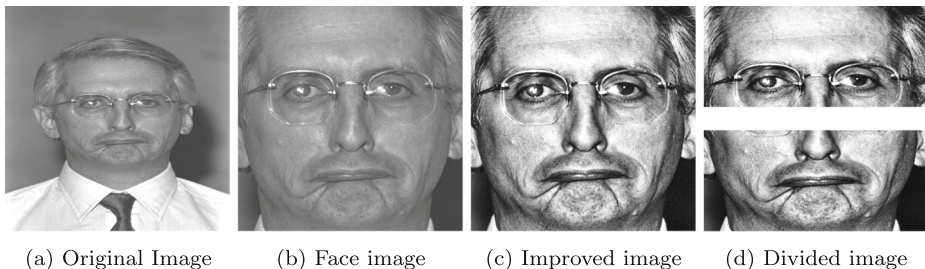


Fig. 2 Four main steps before clustering images

3.2.1 Feature extraction for clustering

The NMF method is used to extract the discriminative features from images. This method is known as an unsupervised approach in which all elements of the decomposed matrix and the obtained matrix factors are forced to be non-negative [16]. This non-negativity makes the resulting matrices easier to inspect and to visualize [64]. Also NMF is able to decompose objects into its parts [43]. For instance, face images can be decomposed into basis elements of the face, including eyes, nose, lips, etc. The NMF method factorizes any given matrix $V \in \mathbb{R}^{m \times n}$ into two matrices, $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{r \times n}$ with an integer $r \ll \min(m, n)$, and both matrices are non-negative (1).

$$V \approx WH \tag{1}$$

Factors W and H are then found by optimizing the problem (2) [28]. The matrix W is called the base matrix and the matrix H is called the coefficient matrix. In other words, according to this method, each matrix V can be written as a linear combination of columns of matrix W which matrix H is its coefficients.

$$\min f(W, H) = \frac{1}{2} \| V - WH \|_F^2 \tag{2}$$

3.2.2 Clustering approach

After extracting the feature vectors from the top and bottom parts of the images, the average feature vectors of each part are calculated. Then the difference between the mean of the top and bottom features is determined. Finally, a threshold is used to cluster the images. In this way, images whose feature difference is greater than or equal to the threshold are in one cluster and the rest of the images (images whose feature difference is less than the threshold level) are in another cluster (3).

$$C(x) = \begin{cases} 1 & \text{if } diff_x < \tau \\ 2 & \text{if } diff_x \geq \tau \end{cases} \tag{3}$$

In (3), $diff_x = mean(f_{top}) - mean(f_{bottom})$, where f_{top} and f_{bottom} correspond to the extracted feature vectors from top and bottom parts respectively. The threshold level (τ) considered in this paper is the median of the differences. By considering the median as the threshold level, the images will be practically divided into two clusters with equal size. At this point, all gallery images are divided into two clusters.

Now with a given probe image y , the cluster related to that person must be specified. To do this, the average of each cluster is considered as the centroid of that cluster. Features of

the probe image, like the gallery image, is extracted by the NMF method after being divided into upper and lower parts. Then the distance of the test image with the centroids is calculated and the nearest cluster is determined. If the distance of the test image from the nearest cluster is larger than the threshold (τ'), the cluster of the nearest centroid is announced as the probe image's cluster. Otherwise, the closest image of the gallery is specified and its cluster is announced as the probe image's cluster (4).

$$C(y) = \begin{cases} C(c) & \text{if } d_{y,c} \geq \tau' \\ C(\arg \min(D)) & \text{if } d_{y,c} < \tau' \end{cases} \quad (4)$$

In (4), D is the distance matrix calculated by the distance measure between the probe image and all the images in the gallery. c is the centroid of the closest cluster to the probe image y , and $d_{y,c}$ is the distance between y and c . τ' is the threshold level that specifies the way that we must use to identify the test image's cluster. This parameter was selected by trial and error.

3.3 Face recognition

After identifying the new image's cluster, its identity must be determined from the faces within the same cluster. At this stage, new features must be extracted to fit the members of the cluster. To do this, two methods of feature extraction, including NMF and FREAK descriptors will be used. Shi-Tomasi's method [52] is used to detect the points of interest.

3.3.1 Feature extraction for recognising faces

Shi-Tomasi's method (also known as Min-Eigen) is one of the interest point detectors. In fact, it's an improvement over the traditional Harris's method for detecting corner points. For detecting corners, the Sum of Squared Differences (SSD) between any region in the image and patches shifted in some directions should be computed. Then a corner response function (R) along with a threshold indicates whether it's a corner region or not [50].

Harris and Stephens [19] computed an approximation to the second derivation of the SSD that is shown in (5). Since the type of the regions in images can be determined by the amount of changes in directions around it, some inferences can be made based on the value of the eigenvalues of H (5):

1. If there is no change in any directions, then it's a flat region ($\lambda_1 \approx 0, \lambda_2 \approx 0$).
2. If there are some significant changes in one direction, then it's an edge region ($\lambda_1 \approx 0, \lambda_2$ has large value).
3. If there are some significant changes in all directions, then it's a corner region (λ_1 and λ_2 both have large values).

$$H = \begin{bmatrix} \widehat{I_x^2} & \widehat{I_x I_y} \\ \widehat{I_x I_y} & \widehat{I_y^2} \end{bmatrix}, \quad (5)$$

where I_x and I_y are the partial derivatives of image I , and $\widehat{}$ denotes averaging (i.e. summation over the patch). Then Harris and Stephens used (6) to compute the corner response. Shi and Tomasi [52] then conclude that it is better to use the minimum of the eigenvalues as the corner response (7). Shi-Tomasi's method has been used in some object tracking [33] and action recognition [63] applications.

$$R = |H| - k(\text{trace}(H))^2 \quad (6)$$

$$R = \min(\lambda_1, \lambda_2) \quad (7)$$

After detecting key-points (corner points), descriptors around those key-points should be calculated. We use FREAK descriptor to extract binary features from Shi-Tomasi's key-points. FREAK is inspired by the human visual system and precisely the retina. Based on FREAK a retinal sampling grid is used which is also circular and have high density in the center. Each circle represents the standard deviation of a Gaussian kernel that applied to the corresponding sampling points. It constructs descriptor F that is a binary string formed by a sequence of one-bit Difference of Gaussians (DoG):

$$F = \sum_{0 \leq a < N} 2^a T(P_a), \quad (8)$$

where

$$T(P_a) = \begin{cases} 1 & \text{if } (I(P_a^{r1}) - I(P_a^{r2})) > 0 \\ 0 & \text{if } (I(P_a^{r1}) - I(P_a^{r2})) \leq 0 \end{cases} \quad (9)$$

and $I(P_a^{r1})$ denotes the smoothed intensity of the first receptive field of the pair P_a . Then a greedy algorithm is used to select best pairs to find uncorrelated tests with high variance. Finally, they are normalized to make descriptors robust to orientation. FREAK descriptor's computation is faster than the previous methods such as SIFT and SURF [3].

NMF features are extracted exactly like the clustering step. Then corner points are detected using Shi-Tomasi's method, and FREAK descriptors are calculated on detected key-points. NMF features are extracted globally from the image, while FREAK descriptors are extracted locally. Global features describe the image as a whole and can be interpreted as a particular property of the image involving all pixels; While, the local features describe individual pixels with their neighbors, and in some approaches, they detect keypoints and describe the regions around these keypoints [21]. Combining these two features makes the proposed method more powerful. Because by combining these two features, both the generality of the image and the details of the image are given importance.

3.3.2 Matching faces

After extracting the features, the image distances matrix is calculated for the two types of features. Finally, the total distance matrix is calculated by combining the distance matrices of the NMF features and FREAK descriptors. In principle, to obtain the total distance matrix, we must calculate the average of the two matrices. In the proposed method, we calculate the weighted average to make the effect of the NMF features different from FREAK descriptors (10). Weight values (α , β) are calculated by trial and error. Then, matching the same images is done by applying the 1-nearest neighbor method with the City-Block distance measure.

$$D_{total} = \frac{\alpha \times D_{NMF} + \beta \times D_{FREAK}}{\alpha + \beta} \quad (10)$$

4 Experiments

In this section, we will discuss the results of experiments performed to evaluate the proposed method. To perform the experiments and to show the robustness of the proposed method to large datasets, the well-known FERET [47] dataset was used as well as an extended image set consisting of images collected from several standard datasets.

The FERET dataset was collected from August 1993 to July 1996 in 15 sessions from 1,199 people. This dataset has a total of 14,126 images that each person can have multiple

images. Images belong to each subject were categorized into frontal, half, quarter, profile, rotated, and duplicate images. In this article, 990 images from the frontal set (fa) are used for training and their corresponding fb images for testing.

Also, an extended image set is collected consisting of frontal images from several standard datasets. These datasets are FERET, MUCT [39], Psychological Image Collection at Stirling (PICS),¹ FEI [59], Face94.² Their main difference is in the brightness or change of expressions. In total, the number of collected subjects reaches 1684 people (Fig. 3).

To perform the experiments, the images are first converted to grayscale and the face image is extracted using the Viola-Jones [62] algorithm. We also examine another face detection method that is called MTCNN [73]. The results shown in Table 1 indicate that Viola-Jones method works better than MTCNN for face image extraction.

Face images are then enhanced using histogram equalization and CLAHE. To extract the NMF features, images are downsampled to size 20×20 pixel. 120×120 pixel images have been used to detect corner points and extract FREAK descriptors.

The first experiment examines the performance of the proposed clustering approach using F-measure (F1-score). F-measure is known as the harmonic mean of precision and recall. F1-score ranges from [0, 1], where 0 indicates the worst and 1 is the best [34].

We compared our clustering method with some common clustering methods including K-means, K-medoids, Agglomerative Hierarchical Clustering (Agglomerative Nesting - AGNES), DBScan, and Spectral Clustering. However, DBScan method failed to cluster images into separate clusters, so this method is not mentioned in the results.

According to Fig. 4, the proposed method performs better than the other clustering methods in terms of both F-measure and final recognition accuracy. AGNES and Spectral Clustering methods also have higher recognition rate, but their clusters are not balanced in terms of cluster size. The balance of the clusters is evaluated using the ratio of cluster members according to (11). For example they split images into two clusters with the size of 20 and 970, while our method guarantees that final clusters will be in the same size. Balanced clustering tends to avoid forming outlier clusters, and thus has beneficial regularizing effect [74].

$$R = \frac{|\text{members of the small cluster}|}{|\text{members of the large cluster}|} \quad (11)$$

Another experiment was performed to investigate the effect of the clustering steps that applied to the dataset. For this purpose, we increased the number of clustering hierarchies. In this case, at each hierarchy, each cluster is further divided into two clusters. For example, if there are two hierarchies of clustering, in the first hierarchy we will have two clusters and in the second hierarchy, we will have two more clusters for each cluster; So the final number of clusters will be four. In this way, a binary tree of clusters can be formed. The structure of the clustering hierarchies is shown in Fig. 5. The results of this experiment are shown in Fig. 6. As can be seen, the final accuracy decreases as the clustering steps increases. Hence the best accuracy is achieved when only one clustering hierarchy is performed.

According to the proposed method, the images were first divided into upper and lower parts. Then we can extract features separately from both parts. We consider two cases to investigate the effect of dividing the images into two parts in the clustering step. In the first case, as described in the proposed method, the images were divided into two parts, top and bottom, and the NMF features were extracted from each part separately. In the second case,

¹ Available on <http://pics.stir.ac.uk>.

² Available on <https://cswww.essex.ac.uk/mv/allfaces/faces94.html>.

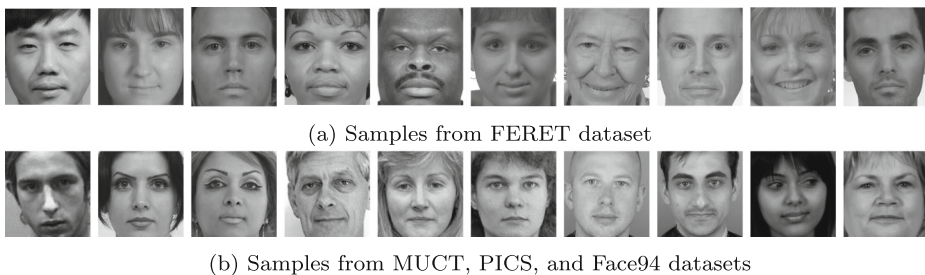


Fig. 3 We evaluated our method using FERET [47] dataset, and an extended image set including FERET [47], MUCT [39], PICS, FEI [59], and Face94 images

the NMF features were extracted from the whole image. Then, the feature vector is divided into two equal parts so that we can calculate the difference between the averages of the two parts.

The results of this experiment are reported in Table 2. These results show that dividing images into two parts and extracting features separately from both parts gives us a much better performance. The rank value in this experiment was obtained by trial and error.

In another experiment, the effect of the different sizes of dataset is evaluated. In this experiment the dataset size starts with a number of 100 images. These images are randomly selected from all the images of the dataset, and the final accuracy is the average accuracy of multiple runs of the proposed method. The results of this experiment are shown in Fig. 7. The proposed method performs well even for large sets of images. As can be seen, as the dataset size increases, the final accuracy decreases at a very small rate.

The experiments were performed on a personal computer with Intel Corei5, 3.2 GHz CPU, and 12 GB RAM. To examine the computational complexity of the proposed method, it was tested with different dataset sizes. The results are reported in Fig. 8. In Fig. 8a, the running time is required for the proposed method to learn the gallery set of the dataset, is shown. In Fig. 8b, the running time is required for an unseen test image to be compared and identified by the proposed method, is recorded. Both running times increase linearly with the increase of the dataset size. It can be said that the computational complexity of the proposed method is approximately in order of N , i.e. ($\mathcal{O}(N)$), which N represents the number of images in the dataset.

We compare the performance of our method with some other existing methods that were described in Section 2. These methods mainly focus on solving dataset size and SSPP problems. The disadvantage of the methods proposed to solve the problem of large dataset size is the need for a large number of samples for training. On the other hand, the disadvantage of the methods proposed to solve the problem of SSPP is the weak performance on the large datasets.

SPCA+ [72], ML-EM [46], FLVD-SVD [48], DMMA [37], Fast-DMMA [35], LU [24], LPP [45], and NMF [42] methods use handcrafted features, while TDL [71] and KCFT [40] methods use transfer learning models to extract features. However, they were evaluated on

Table 1 Comparing the performance of two face detection methods (bold number shows the best result)

Face detection method	Accuracy (%)
Viola-Jones [62]	98.36
MTCNN [73]	96.17

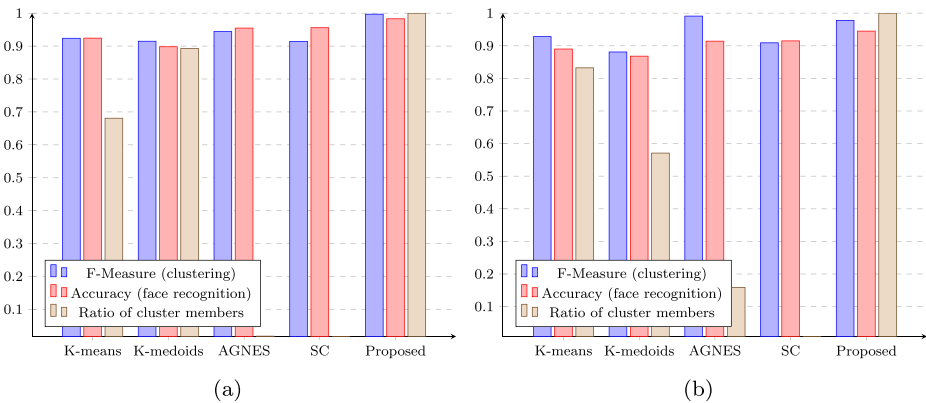


Fig. 4 Performance of the clustering methods on a) FERET dataset, b) extended dataset

small image sets. The accuracy of recognition on the FERET dataset achieved by the mentioned methods is listed in Table 3. The results show that the proposed method outperforms previous methods. While these methods just focus on one problem, the proposed method considers both of the dataset size and SSPP problems simultaneously.

To investigate further with more images in the dataset, we implemented Nikan’s method (NMF) [42], which is the latest published method in this field, and evaluated it on the extended dataset. The result of this evaluation is shown in Table 4. The results indicate that our proposed method can perform better than the NMF method.

One of the important parameters in the proposed method is the rank of the NMF method (r). The value of rank should be an integer that is far less than the dimensions of the initial matrices [42]. For example, if the input matrix is 400×990 (400 is for flattened images that are 20×20 pixel, and 990 is for the number of images in the dataset), then the rank value should be $r \ll \min(400, 990)$. In the proposed method, the value of rank was obtained by

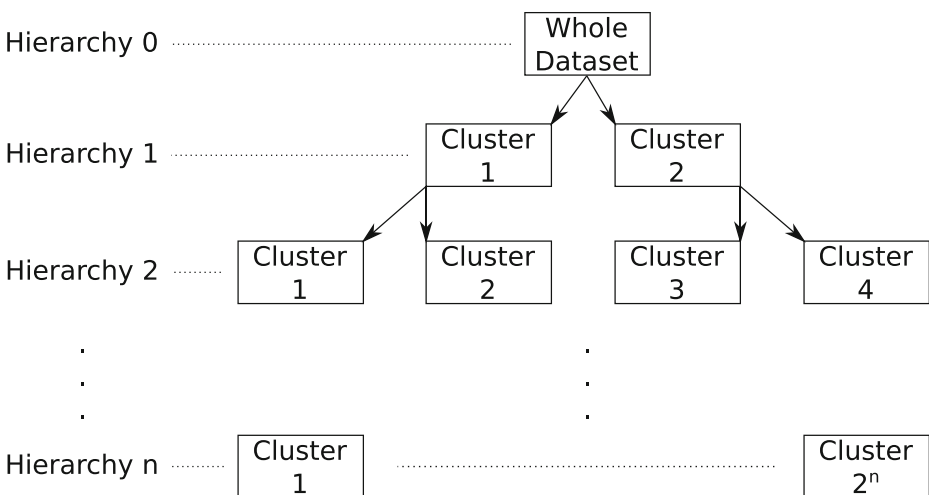


Fig. 5 Hierarchies of the proposed clustering approach

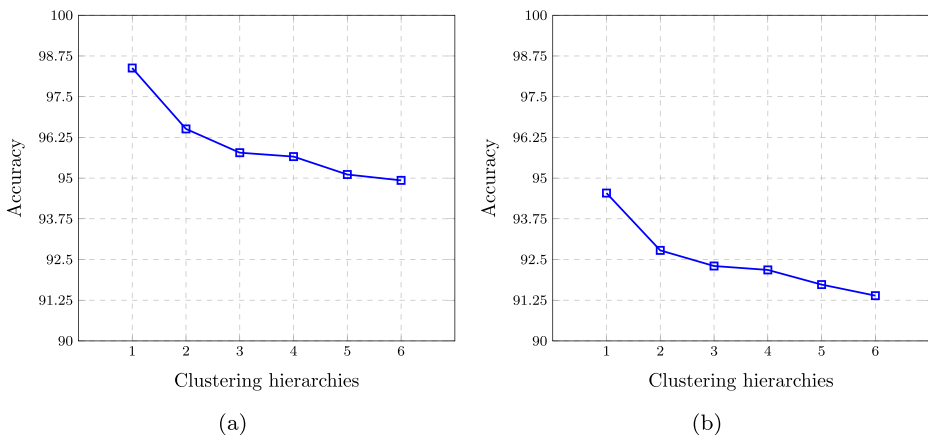


Fig. 6 Effect of clustering hierarchies, evaluated on a) FERET dataset, b) extended dataset

Table 2 Effect of dividing images into two parts in clustering step (bold number shows the best result)

Case	Accuracy (%)
Without dividing images into two parts	73.84
With dividing images into two parts	98.36

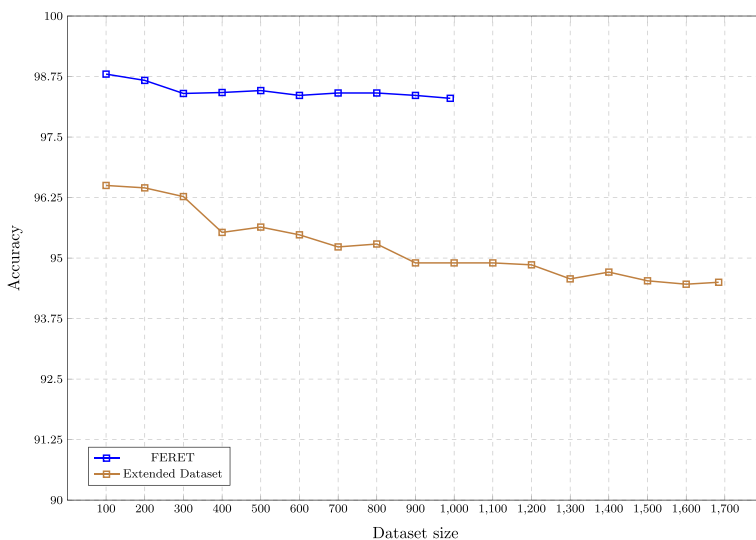


Fig. 7 Effect of the number of images in the dataset on the performance of the proposed method

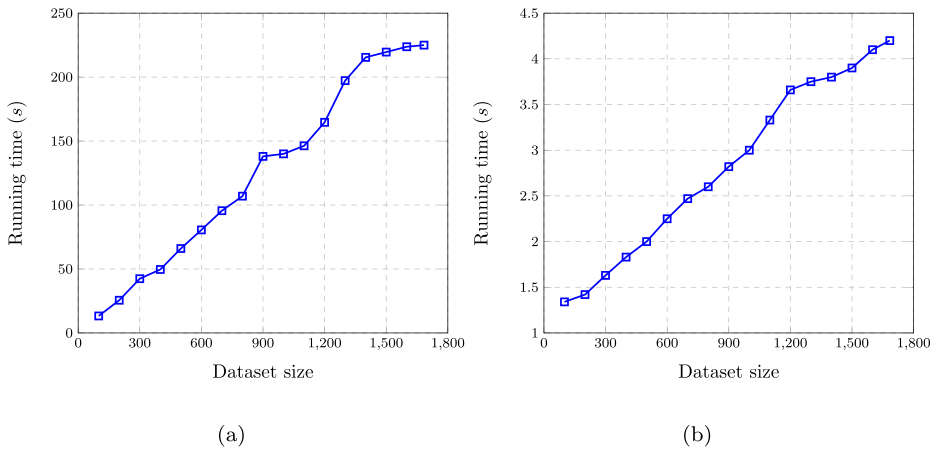


Fig. 8 Required running time for a) training step, b) test an unseen image

trial and error. It can be seen in Fig. 9 that for a number of 990 images from FERET dataset, the rank value of 55 gives the best result.

Since one of the most important steps in this method is to select the nearest image from the gallery collection, an appropriate distance measure must be selected. For this purpose, experiments with several distance measures, including Euclidean, Correlation, Cosine, City-block, and Minkowski were performed. The results of this experiment are known in Table 5. According to the results, the City-block measure was better than other measures to show the same-class images close to each other and different-class images apart.

In another experiment, to investigate the effect of various parts of the proposed method including clustering and use of the two NMF and FREAK descriptors, we applied these factors separately. For example, to see the effect of using NMF features, we applied only the NMF method on the images set; Or to investigate the performance of the proposed method when not using the clustering step, we just applied NMF and FREAK on the images set without clustering. The results are shown in Fig. 10.

Table 3 Comparison of the proposed method with the existing methods on FERET dataset (bold number shows the best result)

Method	Dataset size	Samples per subject	Accuracy (%)
SPCA+ [72]	200	1	87.49
ML-EM [46]	695	1	95
FLDA-SVD [48]	200	1	90.5
DMMA [37]	200	1	93
Fast-DMMA [35]	200	1	91
LU [24]	490	10	91.30
LPP [45]	200	1	89
TDL [71]	200	1	93.9
KCFT [40]	200	1	93.16
RBF+DT [17]	350	1	93.85
NMF [42]	990	1	92.73
Proposed	990	1	98.36

Table 4 Comparison of the proposed method with the existing methods on the Extended dataset (bold number shows the best result)

Method	Dataset size	Samples per subject	Accuracy (%)
NMF	1684	1	89.89
Proposed	1684	1	94.54

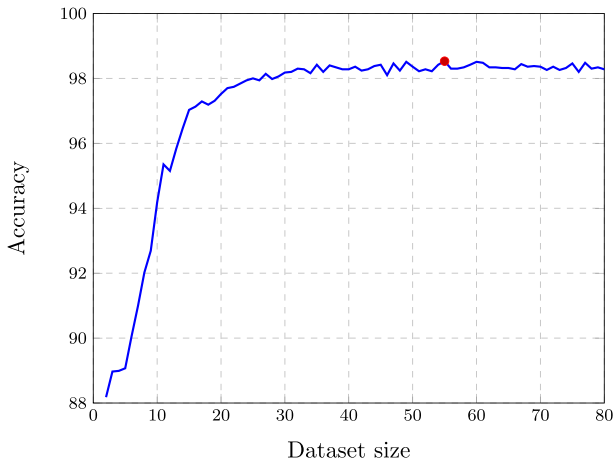


Fig. 9 Effect of the parameter r (rank in NMF) in recognition rate

Table 5 Performance of the proposed method using different distance measures

Distance measure	Accuracy (%)
Euclidean	97.43
Correlation	97.58
Cosine	97.41
City-Block	98.36
Minkowski	97.47

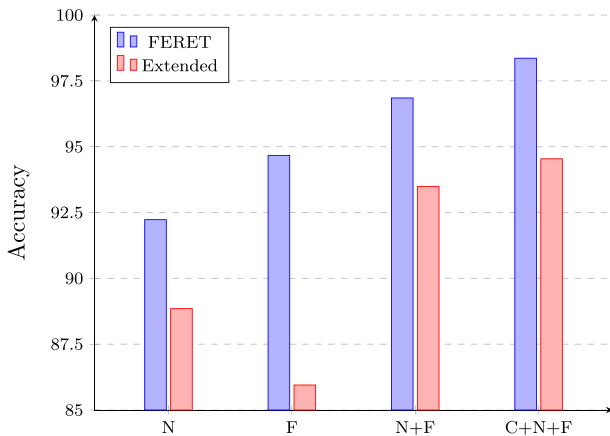


Fig. 10 Effects of various factors on the recognition rate of the proposed method. The letters N, F, and C stand for NMF, FREAK, and Clustering factors

As can be seen, using only NMF or FREAK descriptors has poor performance. But using a combination of these two features has a significant improvement in the final performance. Finally, we added the clustering step to the proposed method, which provides the best performance.

For further comparison, we examined several related works, including Choi's method [12], Gutta's method [17], and Bai's method [5], which use hierarchical approaches and hybrid features for image or face applications. Most of these methods were not mentioned in Table 3. The proposed method uses a balanced clustering approach to create a hierarchy of the dataset. It is guaranteed that the created clusters are always balanced, while the mentioned methods do not generate balanced clusters.

The proposed method uses one sample per subject for training while the mentioned methods use several samples for training. The number of subjects used to evaluate the proposed method is more than a thousand, while most of the previous methods were only able to evaluate a limited number of subjects. Gutta's method [17] and Choi's method [12] use 350 subjects, Bai's method [5] uses 70 subjects to evaluate.

5 Conclusion and future works

In this paper, a hierarchical face recognition method for large datasets is presented. The proposed method also considers the SSPP problem in face recognition. We use a simple method for clustering images, which has performed better than existing clustering methods. Within each step of clustering (hierarchy), images are divided into two clusters. Also, the proposed clustering approach guarantees that the clusters in each steps are balanced and equal. Then two sets of discriminative features including NMF and FREAK descriptors are extracted from faces within each cluster. Finally, the identity of the test image is determined using the nearest neighbor method with the City-Block distance measure.

Previous methods focused on either solving large dataset size or the SSPP problems, whilst the proposed method can consider these two problems simultaneously. Experiments show that the computational complexity of the proposed method is in order of N , which means it can be applied to larger datasets without getting computational problems. Since the employed images in this study have various expressions, experimental results indicate robustness of the proposed method to the expression changes. Other experiments conducted in this research indicate that the proposed method performs better than the previous methods.

The images used in this method are mostly frontal. One of the limitations of the proposed method is the use of images with no variation in pose. Therefore, research can be continued to eventually make the proposed method invariant to pose changes. In this case, the proposed method can be tested on other real-world large-scale datasets. The focus of this research is mainly on facial expressions, and not images with cosmetic and aging effects. Hence, if the given image has any of such effects, the final performance will be affected. This issue will be considered in future works. In addition, research can be continued to improve the clustering method so that by applying several clustering hierarchies, the dataset can be divided into several clusters with a much smaller number of members.

References

1. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. In: Computer vision - ECCV 2004, vol 3021. Springer Berlin Heidelberg, pp 469–481

2. Al-Obaydy WNI, Suandi SA (2020) Open-set face recognition in video surveillance: a survey. In: Kasruddin Nasir AN, Ahmad MA, Najib MS, Abdul Wahab Y, Othman NA, Abd Ghani NM, Irawan A, Khatun S, Raja Ismail RMT, Saari MM, Daud MR, Mohd Faudzi AA (eds) InECCE2019. Springer Singapore, Singapore, pp 425–436
3. Alahi A, Ortiz R, Vanderghyest P (2012) FREAK: fast retina keypoint. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, Providence, pp 510–517
4. Annalakshmi M, Roomi SMM, Naveedh AS (2019) A hybrid technique for gender classification with SLBP and HOG features. *Clust Comput* 22(S1):11–20
5. Bai G, Zhu Y, Ding Z (2008) A hierarchical face recognition method based on local binary pattern. In: 2008 Congress on image and signal processing. IEEE, Sanya, pp 610–614
6. Bartlett M, Movellan J, Sejnowski T (2002) Face recognition by independent component analysis. *IEEE Trans Neural Netw* 13(6):1450–1464
7. Bastanfard A, Takahashi H, Nakajima M (2004) Toward e-appearance of human face and hair by age, expression and rejuvenation. In: 2004 International conference on cyberworlds. IEEE, Tokyo, pp 306–311
8. Biglari M, Neycharan J, Mirzaei F, Soleimani A (2014) Robust feature extraction using gabor filter, lbp and lda for face recognition
9. Chen Y, Zhao Y (2006) Face recognition using DCT and hierarchical RBF model. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Corchado E, Yin H, Botti V, Fyfe C (eds) Intelligent data engineering and automated learning – IDEAL 2006, vol 4224. Springer Berlin, Heidelberg, Berlin, pp 355–362. Series title: lecture notes in computer science
10. Chen S, Liu J, Zhou ZH (2004) Making FLDA applicable to face recognition with one sample per person. *Pattern Recogn* 37(7):1553–1555
11. Cheng Z, Zhu X, Gong S (2020) Face re-identification challenge: are face recognition models good enough? *Pattern Recogn* 107:107422
12. Choi SE, Lee YJ, Lee SJ, Park KR, Kim J (2011) Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recogn* 44(6):1262–1281
13. Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
14. Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)
15. Du G, Su F, Cai A (2009) Face recognition using SURF features, p 749628. Yichang, China
16. Duong VH, Bui MQ, Wang JC (2019) Matrix factorization on complex domain for face recognition. In: Luigi Mazzeo P, Ramakrishnan S, Spagnolo P (eds) Visual object tracking with deep neural networks. IntechOpen
17. Gutta S, Wechsler H (1996) Face recognition using hybrid classifier systems. In: Proceedings of international conference on neural networks (ICNN'96), vol 2. IEEE, Washington, pp 1017–1022
18. Hao L, Yu W (2019) Face image set recognition based on improved HOG-NMF and convolutional neural networks. *Progress Hum Comput Interact* 2(1)
19. Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings of the Alvey vision conference. Alvey Vision Club, pp 23.1–23.6
20. Hassaballah M, Aly S (2015) Face recognition: challenges, achievements and future directions. *IET Comput Vis* 9(4):614–626
21. Hassaballah M, Awad AI (2016) Detection and description of image features: an introduction. Series title: studies in computational intelligence, vol 630. Springer International Publishing, Cham, pp 1–8
22. Hassaballah M, Awad AI (eds) (2020) Deep learning in computer vision: principles and applications. Digital imaging and computer vision, 1st edn. CRC Press/Taylor and Francis, Boca Raton
23. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. arXiv:1512.03385
24. Hu C, Lu X (2015) Single sample face recognition via lower-upper decomposition. In: 2015 10th asian control conference (ASCC). IEEE, Kota Kinabalu, pp 1–6
25. Ji HK, Sun QS, Ji ZX, Yuan YH, Zhang GQ (2017) Collaborative probabilistic labels for face recognition from single sample per person. *Pattern Recogn* 62:125–134
26. Kan M, Shan S, Su Y, Xu D, Chen X (2013) Adaptive discriminant learning for face recognition. *Pattern Recogn* 46(9):2497–2509
27. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev*. arXiv:1901.06032
28. Kim J, Park H (2011) Fast nonnegative matrix factorization: an active-set-like method and comparisons. *SIAM J Sci Comput* 33(6):3261–3281
29. Kortli Y, Jridi M, Falou AA, Atri M (2020) Face recognition systems: a survey. *Sensors* 20(2):342

30. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
31. Lee DD, Seung HS (2000) Algorithms for non-negative matrix factorization. *Neural Inf Proces Syst* 13:7
32. Li SZ, Jain AK (eds) (2011) *Handbook of face recognition*. Springer, London
33. Li H, Doermann D, Kia O (2000) Automatic text detection and tracking in digital video. *IEEE Trans Image Process* 9(1):147–156
34. Li Y, Hu H, Zhu Z, Zhou G (2020) SCANet: sensor-based continuous authentication with two-stream convolutional neural networks. *ACM Transactions on Sensor Networks* 16(3):1–27
35. Liu HH, Hsu SC, Huang CL (2014) Single-sample-per-person-based face recognition using fast discriminative multi-manifold analysis. In: *Signal and information processing association annual summit and conference (APSIPA), 2014 Asia-Pacific*. IEEE, Chiang Mai, pp 1–9
36. Lu J, Plataniotis KN (2002) Boosting face recognition on a large-scale database. In: *Proceedings. international conference on image processing, vol 2, pp II–II*
37. Lu J, Tan YP, Wang G (2013) Discriminative multi-manifold analysis for face recognition from a single training sample per person. *IEEE Trans Pattern Anal Mach Intel* 35(1):39–51
38. Martinez A (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans Pattern Anal Mach Intel* 24(6):748–763
39. Milborrow S, Morkel J, Nicolls F (2010) The MUCT landmarked face database. *Pattern Recognition Association of South Africa* 201(0)
40. Min R, Xu S, Cui Z (2019) Single-sample face recognition based on feature expansion. *IEEE Access* 7:45219–45229
41. Moussa M, Hmila M, Douik A (2018) A novel face recognition approach based on genetic algorithm optimization. *Studies in Informatics and Control* 27(1)
42. Nikan F, Hassanpour H (2020) Face recognition using non-negative matrix factorization with a single sample per person in a large database. *Multimed Tools Appl* 79:28265–28276
43. Nikolaus R (2007) Learning the parts of objects using non-negative matrix factorization (nmf). MMer Team
44. Otto C, Wang D, Jain AK (2016) Clustering millions of faces by identity. arXiv:1604.00989 [cs]
45. Pan J, Wang XS, Cheng YH (2016) Single-sample face recognition based on LPP feature transfer. *IEEE Access* 4:2873–2884
46. Perronnin F, Dugelay JL (2005) Clustering face images with application to image retrieval in large databases, p 256. Orlando, Florida, USA
47. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (1999) The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intel* 22(10):1090–1104
48. Qx G, Zhang L, Zhang D (2008) Face recognition using FLDA with single training image per person. *Appl Math Comput* 205(2):726–734
49. Reza AM (2004) Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology* 38(1):35–44
50. Rosten E, Drummond T (2006) Machine learning for high-speed corner detection. In: *Computer vision – ECCV 2006, vol 3951*. Springer Berlin, Heidelberg, pp 430–443
51. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: *2015 IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE, Boston, pp 815–823
52. Shi J, Tomasi C (1994) Good features to track. In: *Proceedings of IEEE conference on computer vision and pattern recognition CVPR-94*. IEEE Comput. Soc. Press, Seattle, pp 593–600
53. Simonyan K, Parkhi O, Vedaldi A, Zisserman A (2013) Fisher vector faces in the wild. In: *Proceedings of the British machine vision conference 2013*. British Machine Vision Association, Bristol, pp 8.1–8.11
54. Sireesha V, Latha L, Anuradha A, Kumar NP (2019) Histogram equalization techniques in image enhancement. *J Info Tech Softw Eng* 09((01)):254
55. Sun Y, Wang X, Tang X (2014) Deep learning face representation from predicting 10,000 classes. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*
56. Sun L, Zhao C, Yan Z, Liu P, Duckett T, Stolkin R (2019) A novel weakly-supervised approach for RGB-D-based nuclear waste object detection. *IEEE Sensors J* 19(9):3487–3500
57. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*
58. Taskiran M, Kahraman N, Erdem CE (2020) Face recognition: past, present and future (a review). *Digit Signal Process* 106:102809

59. Thomaz CE, Giraldi GA (2010) A new ranking method for principal components analysis and its application to face image analysis. *Image Vis Comput* 28(6):902–913
60. Tuncer T, Dogan S, Abdar M, Plawiak P (2020) A novel facial image recognition method based on perceptual hash using quintet triple binary pattern. *Multimedia Tools and Applications*
61. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86
62. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol 1. IEEE Comput. Soc, Kauai, pp 1–511–I–518
63. Wang H, Schmid C (2013) Action recognition with improved trajectories. In: 2013 IEEE international conference on computer vision. IEEE, Sydney, pp 3551–3558
64. Wild S, Curry J, Dougherty A (2004) Improving non-negative matrix factorizations through structured initialization. *Pattern Recogn* 37(11):2217–2232
65. Wu B, Liu Z, Yuan Z, Sun G, Wu C (2017) Reducing overfitting in deep convolutional neural networks using redundancy regularizer. In: Lintas A, Rovetta S, Verschure PF, Villa AE (eds) Artificial neural networks and machine learning – ICANN 2017, vol 10614. Springer International Publishing, Cham, pp 49–55
66. Xi M (2015) Local binary patten network: a deep learning approach for face recognition. In: 2016 IEEE international conference on Image processing (ICIP), pp 3224–3228
67. Xin M, Zhou Y, Yan J (2015) Single sample face recognition using lgbp and locality preserving discriminant analysis. *Applied Mathematics & Information Sciences* 9:353–360
68. Yang H, Wang XA (2016) Cascade classifier for face detection. *J Algorithms Comput Tech* 10(3):187–197
69. Yang M, Van L, Zhang L (2013) Sparse variation dictionary learning for face recognition with a single training sample per person. In: 2013 IEEE international conference on computer vision. IEEE, Sydney, pp 689–696
70. Yang M, Wen W, Wang X, Shen L, Gao G (2020) Adaptive convolution local and global learning for class-level joint representation of facial recognition with a single sample per data subject. *IEEE Trans Info Forensics Secur* 15:2469–2484
71. Zeng J, Zhao X, Gan J, Mai C, Zhai Y, Wang F (2018) Deep convolutional neural network used in single sample per person face recognition. *Comput Intell Neurosci* 2018:1–11
72. Zhang D, Chen S, Zhou ZH (2005) A new face recognition method based on SVD perturbation for single example image per person. *Appl Math Comput* 163(2):895–907
73. Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process Lett* 23(10):1499–1503
74. Zhong S, Ghosh J (2003) Model-based clustering with soft balancing. In: The 3rd IEEE international conference on data mining (ICDM), pp 459–466
75. Zhou H, Yuan Y, Shi C (2009) Object tracking using SIFT features and mean shift. *Comput Vis Image Underst* 113(3):345–352
76. Zhou E, Cao Z, Yin Q (2015) Naive-deep face recognition: touching the limit of LFW benchmark or not? arXiv:1501.04690

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.