Multimedia Tools and Applications (2020) 79:27891–27922
https://doi.org/10.1007/s11042-020-09261-2

A review on face recognition systems: recent approaches and challenges

Muhtahir O. Oloyede[1,2] • Gerhard P. Hancke[2] • Hermanus C. Myburgh[2]

Received: 8 August 2019 / Revised: 16 April 2020 / Accepted: 24 June 2020 /
Published online: 30 July 2020

Abstract

Face recognition is an efficient technique and one of the most preferred biometric modalities for the identification and verification of individuals as compared to voice, fingerprint, iris, retina eye scan, gait, ear and hand geometry. This has over the years necessitated researchers in both the academia and industry to come up with several face recognition techniques making it one of the most studied research area in computer vision. A major reason why it remains a fast-growing research lies in its application in unconstrained environments, where most existing techniques do not perform optimally. Such conditions include pose, illumination, ageing, occlusion, expression, plastic surgery and low resolution. In this paper, a critical review on the different issues of face recognition systems are presented, and different approaches to solving these issues are analyzed by presenting existing techniques that have been proposed in the literature. Furthermore, the major and challenging face datasets that consist of the different facial constraints which depict real-life scenarios are also discussed stating the shortcomings associated with them. Also, recognition performance on the different datasets by researchers are also reported. The paper is concluded, and directions for future works are highlighted.

Keywords Face recognition · Biometrics · Techniques · Uncontrolled environment · Face dataset

1 Introduction

Face recognition (FR) has over recent years been an active research area due to the various applications it can be applied, such as border security, surveillance, law enforcement and access control. Recently, other applications involved with the FR system include computer

✉ Muhtahir O. Oloyede
tahir.oloyede@gmail.com

[1] Department of Information and Communication Science, University of Ilorin, Ilorin, Nigeria

[2] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

graphics, neural networks, and psychology as it is more of a multidisciplinary interest lately. Similar to other biometric systems, as shown in Fig. 1, the stages involved in the FR process are face detection, pre-processing of the face image, extraction of facial features and lastly feature classification [65]. The first stage which is the detection of the face is the process where the system is made to verify the presence of a face in an image or video. After detection of the face, the image is pre-processed to acquire the region of interest and also improve the image quality. Normalization is a type of pre-processing technique where the face images of different scales are transformed and mapped into the same scale. Face alignment is another pre-processing approach which is the process of localizing the fiducial points such as mouth, eyes chin and nose. This approach is seen to improve face recognition system, though it remains a challenge and an open problem in unconstrained environments [39]. Image enhancement is also a pre-processing approach that is overlooked in the literature. Its primary objective in this regard is to come up with an enhanced face image from the original, which is supposed to improve the overall performance of the face recognition system [42]. Feature extraction is the next stage of the face recognition system model and its purpose is to simplify the number of resources that describes a large set of data. Also, the extraction of unique features is done to minimize noise and irrelevant information present in the original face image, and a feature vector that is sufficient in describing the face from the face image is extracted. Many feature extraction methods have been proposed in the literature, however, selecting the right features for various face recognition systems in the unconstrained environment remains a challenge [20]. Most common classifiers used in the field of face recognition are the minimum distance classifier, nearest neighbor classifier, and k- nearest neighbor classifier [6]. The minimum distance classifier places the label of a testing sample as the class whose mean it is associated.The nearest neighbor classifier places testing samples into the class that its nearest neighbor is associated to, while the k – nearest neighbor classifier places the testing samples with the class that has the nearest neighbor by first searching for the k-nearest neighbor [63].In recent times, machine learning algorithms have been the most preferred to use classification techniques such as the convolutional neural network (CNN). The feature classification stage that leads to the recognition of face images involves both identification and authentication of facial images. Identification compares a face image with other face images to be able to come up with an identity of the face among several possibilities while authentication is when a face is compared with another to approve the requested identity [9]. In both scenarios, face images of known individuals are registered in the system known as a gallery. After that, face images which are used as probes can either be the registered or unregistered individuals used for the identification or recognition task.

The existence of face recognition systems over the years has been applied to various applications which has encouraged researchers in both academia and industry to come up

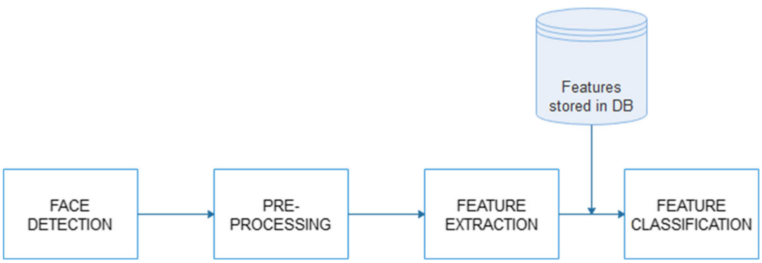

Fig. 1 Flow diagram of the stages involved in a face recognition system

with various approaches and techniques for FR systems. These developed techniques perform optimally under well-controlled conditions. However, they tend to fail in real-life conditions, also known as face recognition in the wild or unconstrained environment. In these situations, face images are captured in a way that makes processing a difficult task, thus leading to the problem that differences of the face of the same individual are higher than that of different individuals, which leads to a reduction in recognition rate. These real-life conditions include occlusion, illumination, facial expression, plastic surgery, low image resolution and ageing. It is expected that a robust face recognition system can perform optimally in the listed real-life scenarios also termed "face recognition in the wild." Hence, face recognition in the unconstrained environment is still seen to be a lingering issue that has not yet been solved. Different surveys of face recognition systems in the unconstrained environment has been carried out by researchers,but limited discussion was done. Ghiass et al.[27]gave an up to date survey on face recognition system but was limited to only infrared face recognition approach. Goyal et al. [28], also gave a review on face recognition systems, however only facial expression as a constraint was considered in their work. A survey on heterogeneous face recognition was presented by Ouyang et al. [69], that described matching face images across different domains. They further discussed benchmark datasets used for evaluation. However, they did not present the state-of-the-art performance on these datasets. Thus, this work is unique in that the major constraints of the face recognition system are discussed and stating why they are still lingering issues that do not enable the system to perform optimally. The survey paper has been presented to target both new and existing researchers in this field. The main contribution of this paper is to critically review the issue of face recognition systems in a real-life scenario or the unconstrained environments, stating why they still need to be researched on. An up to date state-of-the-art performance on the various face datasets containing the different constraints is reported. Also, we briefly discuss the various existing techniques of the face recognition system, stating their advantages and disadvantages and also why they do not perform optimally in an unconstrained environment. Description of other sections of this paper includes: Section 2 presents the different issues of face recognition systems, highlighting the different recent approaches to each problem. Section 3 presents a review of the standard face recognition techniques. Section 4 discusses the various face datasets that consist of the major facial constraints, and we highlight some of the shortcomings therein. Also, research works from different researchers showing their results on the different face datasets are reported. A general discussion and areas for future research direction are presented in Section 5 while Section 6 concludes the paper.

2 Face recognition in unconstrained environments

Face recognition is regarded as one of the complex biometric systems in the area of pattern recognition because of its constraints caused by changes in the appearance of facial images [29, 39, 96]. The various face recognition techniques have recorded remarkable success in well-controlled environments. However, these techniques tend to fail as a result of the non-stable structure of face images and also in real-life scenarios as it makes the face images of the same person look different [99]. Till date, most face recognition algorithm have not attained an optimal level of recognition accuracy due to these issues, as face acquisition process can undergo to a wide range of variations [1, 4]. The issues of the face recognition system include illumination, pose variation, expression, plastic surgery, ageing low resolution and most importantly occlusion [81]. This section further describes the different issues which are also seen as the face recognition system in unconstrained environment or face recognition in the wild. A detailed presentation of recent and different approaches to each issue is also presented.

2.1 Illumination

Illumination is referred to as lighting variations where a face image can appear differently due to change in lighting. This change in lighting as shown in Fig. 2 appear greater than the differences amongst individuals, hence causing face recognition systems to fail by misclassifying faces when comparing images [12]. Illumination remains an issue to robust face recognition system that involves extracting illumination invariant features. Illumination has great effects on an image that leads to change in location, shadow shape, and reversal of contrast gradients [108]. It is normal for humans to recognize face images regardless of changes in lighting; however, this task remains an issue for face recognition systems. Earlier research works demonstrated using various image representations such as grey-scale, edge maps, filtered images with Gabor filter showed that changing the sides of lighting conditions results to large image differences as compared to changing the identity of the face [58]. Furthermore, testing or training is also seen to be sensitive under varying illumination conditions [41, 108]. Thus, this has made the face recognition system fail in an illumination scenario and has over the last decade attracted researchers in the field to come up with various algorithms to solve this. Algorithms proposed to handle the issue of illumination are of different approaches. Firstly, the use of image processing techniques that are used to normalize face images with various lighting conditions [12]. In this respect, Gamma intensity correlation, logarithms transform or histogram equalization in the form of enhancement has proved to perform well in different lighting conditions. Secondly, the use of the 3D face model approach has been used for the illumination issue [89]. It is seen that frontal face images with various lighting condition result in an illumination cone which is formed on the subspace. These are quite easy to estimate on low dimensional subspace using generative model applied on training data. However, the 3D face model approach needs more amounts of training samples, and also the source of light might need to be specified, which is not an ideal process for a real-life scenario. The third approach involves processes where features of the parts where the illumination occurs are extracted, which are further passed to the recognition stage [93]. Till date, algorithms have been proposed under these approaches; however, research is still on to achieve optimal performance. Table 1 further shows a summary of recent approaches to the issue of illumination.

In [54], the authors proposed two methods that are based on local histogram specifications to pre-process face images affected by different lighting conditions. The idea behind their approach was to effectively remove both the high and low-frequency sections of illumination present on the face image, and also enhance facial features present within the lower frequency

Fig. 2 Changes in illumination adapted from the Yale face dataset

Table 1 Summary of recent face recognition system approaches to the issue of illumination

Technique	Research work	Face dataset	Comments
Local histogram specification	Liu et al. [54]	Extended Yale	Removal of both high and low frequency sections of illumination present on the face image, and enhance facial features present within the lower frequency section.
Adaptively weighted ULBP_MPHOG and WSRC method with Convolutional neural network	Wang et al. [90]	CMU-PIE	Face images are normalized on which uniform local binary pattern and multiple histograms of oriented features are being extracted in each block. Information entropy is further used to achieve the adaptively weighted ULBP_MHOG features. The convolutional neural network is used for the classification purpose.
Discriminative multi-layer illumination robust feature extraction	Yu et al. [95]	Extended Yale	Large-scale features are sectioned into smaller parts, and weights are assigned to every layer to fine-tune its importance. This assist in taking full advantage of the information in large scale features for recognition.
Sparse error of robust PCA	Luan et al. [57]	AR	The obtained sparse error component displays more important information beneficial for face identification. Two recognition protocols i.e. weighted based method and ratio-based method were utilized to classify face images.

section. Firstly, a high pass filter is applied on the face image to filter the low-frequency illumination, on which local histogram and local histogram statistics is applied on the entire face image and are learned from normal lighting images. In their second approach, it is assumed that the regions contain a higher frequency lighting condition and local histogram statistics knows weak facial features, after that local histogram specification is applied on these sections to reduce high-frequency illumination and enhance delicate facial features. For their experiment and to evaluate the performance of their approach, three public face dataset was utilized, i.e., CMU PIE dataset, CAS-PEARL dataset, and the Extended Yale B dataset. To further confirm the effectiveness of their methods various distance metrics were used such as distance transformation, correlation, cosine, and histogram distance. On the CMU PIE dataset, the recognition rate of 78.15% was achieved using the cosine distance metrics, 78.57% was achieved using correlation, 84.54% and 84.68% using histogram distance and distance transformation respectively. In [90], the authors proposed an adaptively weighted ULBP_MPHOG and WSRC method. Firstly, the face images are normalized on which uniform local binary pattern and multiple histograms of oriented features are being extracted in each block. Information entropy is further used to achieve the adaptively weighted ULBP_MHOG features. Experiments were carried out using the CNN on the ORL, CMU PIE, Yale and Extended Yale face dataset based on the varying number of blocks, type of features and different classifiers. On the ORL dataset, a recognition rate of 90.38% was achieved using 4×4 blocks and six training samples per class. On the Yale dataset, a recognition rate of 87.68% was achieved using 10×10 blocks and 3, 4 and 5 training samples per class. Finally, on the Yale B dataset, the recognition rate shows to be better using 16×16 blocks, achieving a recognition rate of 84.3%. Also, in [95], the authors proposed a discriminative multi-layer illumination robust feature extraction model to attempt the illumination problem. Firstly, the multi-layer large-scale features were decomposed into smaller scale illumination robust features as a linear combination. A weight is then assigned to each layer to fine-tune its relevance, to take full advantage of this relevant information in large-scale features for face recognition. Secondly, a discriminant filter is learned to improve the robustness and discriminative analytical capability of the reconstructed illumination robust face for face recognition under varying lighting conditions. Experiments carried out on the FRGC dataset displayed a satisfactory recognition rate. In [57], authors proposed a sparse error of robust PCA that displays important features effective for facial identification. The facial images are further classified using two recognition protocols i.e. weighted based method and ratio-based method. Experiments carried out showed satisfactory performance of their approach.

2.2 Pose variation

The performance of a face recognition system also tends to fail when there is the presence of pose variation in the input images as shown in Fig. 3. In practices such as passport control, face images are mostly on a frontal view. However, in uncontrolled environments where face images can appear in various angles due to rotation, this can cause the face recognition system to fail [3]. Pose variation refers to recognizing face images presented in a different pose; this might be a simple task for human; it remains an issue for computers most especially computer vision applications such as surveillance [22]. With the increase in the rate of terrorist attacks all around the globe, most airports are equipped with surveillance cameras. Face images of terrorist are captured and stored which are later compared with the face images of genuine travelers as their faces will be scanned. These images of the terrorist must have been captured in various pose at an earlier time making the recognition easy using simple face recognition algorithms. In real-life situations where the face images across different pose are not available in the dataset, the system will not perform optimally as the different pose of the same individual will appear as a non-match [89]. When there is the absence of pose tolerance in a face recognition system, the system becomes non-passive and non-intrusive. Hence, there is the desired need for face recognition systems to work optimally under different poses which may be different from the face image stored in the dataset. A major issue faced by the various face recognition techniques in a pose variation situation is to acquire unique features free from pose variation.

In situations where the gallery images stores face images in a different pose, there will be better performance in recognizing a particular face image even in a different pose [23]. This has been shown in the literature, where different face recognition techniques performed better when multiple face images per individual were stored in the dataset as compared to just one face image [33]. This increase in performance is as a result of the techniques being able to tolerate small pose variations i.e. as there is an increase in the number of gallery images, there is a high probability that the probe pose will be close to an image in the gallery [7]. Adding

Fig. 3 Changes in pose adapted from the ORL dataset

more side view images in the gallery gives more required information of the human face structures leading to better-reconstructed models as compared to the use of single gallery images. Most face recognition datasets contain few numbers of gallery images such as police mug shot images that has a frontal and one side view image. Hence, the requirement of various pose images in the gallery restricts the applicability of the techniques, and the ideal situation is to identify a probe image in a random pose from a single gallery image also present in a random pose. This is seen to be more challenging than a multiple gallery view situation. Various face recognition techniques have been proposed and used on benchmark face datasets that have the presence of the pose variation constraints. Benchmark face datasets that have the pose variation constraints include the FERET face dataset, CMU-PIE face dataset, ORL face dataset, WVU dataset, MIT face dataset, Yale B face dataset. Of these listed datasets, FERET and CMU-PIE have either more face images and pose variations as compared to other face datasets. Performances by the various face recognition techniques on these benchmark datasets have not been optimal. In [106], the authors proposed a local binary pattern-like feature extraction method that adjusts the code rule by Huffman. Also, a strategy based on divide and rule is applied to both the face representation and classification with the purpose of improving recognition performance with different pose positions. Experiments carried out on the CMU PIE dataset showed a recognition rate of 84.85%, while that on FERET showed a recognition rate of 85.17%. In [34], an approach based on using landmarks of the face image and depth warping is proposed for robust cross pose face recognition. Different from existing 3-D reconstruction, their approach uses the spontaneously identified broad facial landmarks to change the computationally expensive 3-D reconstruction procedure. During the process of matching to the probe face image, the registered depth-warped faces in the dataset are rotated to match the orientation of the probe image, where sparse regression is used to identify the individual correctly. Experiments were carried out on the PIE and Multi-PIE dataset showing recognition rate 93.5%% and 89.58% respectively. Also, in [20], the authors proposed a highly effective pose-invariant face recognition system. A dense grid of 3D facial landmarks is projected to each 2D face image that allows extraction of facial features in a pose adaptive way. Secondly, an optimal warp is projected to correct texture distortion by the difference in the pose for a local patch around each landmark. Experiments carried out on public datasets showed satisfactory results in constrained and unconstrained environment. Table 2 further shows summary of recent face recognition approaches to the issue of pose variation.

2.3 Expression

Human beings continually have the ability to display various facial expressions unless if the face image is in a static mode [28]. These expressions are used to represent different emotions and mental states of a particular individual as shown in Fig. 4. Also, the expression issue of face recognition is not only seen as an identity or verification problem but also used for medical applications where a particular facial expression can be linked to a particular ailment [87]. It is found out that 7% of the whole information an individual express is passed through language, speech represents 38%,and facial expression represents 55% [82]. From this, we can deduce that a substantial amount of valuable information can be collected to detect an individual's consciousness and mental activities through expression. Facial expression results in a change of the facial appearance and geometry leading to a reduction of recognition performance. This has motivated researchers to come up with different approaches in trying to solve this issue.

Table 2 Summary of recent face recognition system approaches to the issue of pose variation

Technique	Research work	Face dataset	Comments
Local binary pattern enhanced by divide and rule strategy.	Zhou et al. [106]	FERET	The local binary pattern features change the code rule by Huffman. A divide and rule strategy is applied on the face classification for recognition.
3-D reconstruction procedure based on facial landmarks and sparse regression.	Hsu et al. [34]	Multi-PIE	Broad facial landmarks are used to change the computationally expensive 3-D reconstruction process. For matching purpose, registered depth-warped faces in the dataset are rotated to match the position of the probe image.
Convolutional neural network	Patacchiola et al. [70]	LFW	The fusing of dropout and adaptive gradient increase recognition performance.

In [61], the authors proposed a biological-based disparity energy model that can produce 3D disparity maps suitable for identity and verification recognition with different facial expressions. The disparity energy model utilizes two neuronal populations namely encoding population and higher-level decoding population to estimate local disparities. The encoding population consists of neurons adjusted to a wide range of parameters such as orientations, frequencies and horizontal disparities. The population is trained to learn the population codes of many different disparities. The higher-level decoding population makes a comparison of the local population code at each image position, with all trained codes for estimating local disparity. Experiments done on the BU-3DFE dataset show a recognition rate of 89.33%. In [78], the authors proposed for a pose-invariant face recognition system, an enhanced, modified

Table 3 Summary of recent face recognition system approaches to the issue of expression

Technique	Research work	Face dataset	Comments
3-D disparity maps based on biological disparity model	Martins et al. [61]	BU-3DFE	The disparity energy model utilizes two neuronal populations namely encoding population and higher-level decoding population to estimate local disparities. The encoding population consists of neurons adjusted to a wide range of parameters such as orientations, frequencies and horizontal disparities. The population is trained to learn the population codes of many different disparities. The higher-level decoding population makes a comparison of the local population code at each image position, with all trained codes for estimating local disparity.
Local tenary pattern descriptor with support vector machine	Revina et al. [78]	JAFFE	A decision based unsymmetric trimmed median filter used to eradicate noise from the face image; and a dominant gradient local ternary pattern descriptor for feature extraction with the support vector machine classifier for classification.
A 3-D technique using multiple subject-specific curves	Li et al. [52]	GavabDB	The different curves are joined with appropriate weights at the feature level and used for matching 3D faces with the iterative closest point algorithm.

Table 4 Summary of recent face recognition system approaches to the issue of plastic surgery

Technique	Research work	Face dataset	Comments
Shift-invariant fourier transform based on entropy with support vector machine.	Sable et al. [80]	Plastic surgery	Extracts essential regions and volume of the scale-space structure for which the information rate is known.
Gabor feature representation based on edges	Chude et al. [17]	LFW	The edge information that is dependent on the shape of the essential components of the face are used, to address the difference in texture caused by the plastic surgery.
GIST global descriptor with local binary patterns	Ali et al. [2]	Plastic surgery	LBP features with information such as corners and edges are useful; while the GIST descriptor is for gettingbasic and subordinate description of related dimension.

decision based unsymmetric trimmed median filter used to eradicate noise from the face image; also proposed a dominant gradient local ternary pattern descriptor for feature extraction with the support vector machine classifier for classification. Experimental results on the JAFFE dataset gave a performance recognition accuracy of 88%. Authors in [52], proposed a 3D face recognition system that utilizes multiple subject-specific curves insensitive to intra-subject differences caused by a change in expressions. They took into consideration that most sharp variances present in the facial convex regions are firmly related to the bone structure. Hence the extraction of convex crest curves is carried out. After that, the central profile curve and the horizontal contour curve passing through the nose tip are detected using the exact localization of the nose tip and symmetry plane. Due to their unique strength to change in expression, the different curves are joined with appropriate weights at the feature level and used for matching 3D faces with the iterative closest point algorithm. An experiment carried out on GavabDB, and BU-3DFE 3D public dataset showed a recognition rate of 85.2% and 88.9% respectively. In [56], authors to solve the issue of not having enough data in public dataset for facial expression recognition applied specific pre-processing approaches that can extract only expression specific features from the face image. After that, a convolutional neural network architecture was used as the classifier. Experiments on the JAFFE dataset achieved a

Table 5 Summary of recent face recognition system approaches to the issue of ageing

Technique	Related work	Face dataset	Comment
Identity inference model based on age-subspace learning	Zhou et al. [104]	FGNET	The human identity and ageing variables are modeled using probabilistic linear discriminant analysis. Thereafter, the ageing subspace is trained independently with the appearance-age labels, and the identity subspace is then determined with the expectation-maximization algorithm.
Couple autoencoder networks based on neural networks	Xu et al. [92]	FGNET	Bridged by two shallow neural networks used to fit complex nonlinear ageing and de-ageing process. Non-Linear factor analysis is presented to decompose a face image into an identity feature that is age-invariant.

Table 6 Summary of recent face recognition system approaches to the issue of occlusion

Techniques	Related work	Face dataset	Comments
Sparse representation classification	Zhao et al. [102]	AR	Sparse representation classification is applied to represent the query image and obtain the residuals of each class, where each pixel is estimated as either occluded or not in each class's residual.
Sparse error correction with Markov random field algorithm	Fu et al. [24]	Yale B	Removes occluded sections by locally constrained coding and reduces running time for recognition.
Convolutional neural network fuzzy max-pooling method	Long et al. [55]	AR	A fuzzy max-pooling approach based on the convolutional neural network was proposed, where unreliable local features are suppressed from occluded regions. An average pooling is further used to enhance the robustness by automatically weighting on each subclass.

recognition rate of 89.76%. Table 3 further shows a summary of recent approaches to the issue of expression.

2.4 Plastic surgery

It is observed that face recognition techniques fails to identify individual's faces after the process of plastic surgery. This is a situation where the face image of an individual is entirely changed, thus changing to an entirely different individual [64]. The plastic surgery process on the human face introduces change in skin texture between images of the same individual, thus making the face recognition task difficult [2]. A typical scenario of the plastic surgery and how it can cause a change to the face image is rhytidectomy. This is a procedure where the global appearance of the face is completely changed as shown in Fig. 5. This can enhance the change of aged skin texture to a younger state, thus introducing a change in skin texture. Eye lift, reshaping of the nose, enhancement of the jaw are major changes done on the face to cause a change in facial appearance [17]. Face recognition techniques have been developed such that they are robust to the issue of plastic surgery. The dataset used to evaluate face recognition techniques on the issue of plastic surgery include the plastic surgery face dataset.

In [80], the authors proposed an approach called entropy based volume SIFT for recognizing faces after undergoing plastic surgery. Their approach extracts essential regions and volume of

Table 7 Summary of recent face recognition system approaches to the issue of low resolution

Technique	Related work	Face dataset	Comments
Cluster-based discriminant analysis technique.	Chu et al. [16]	SCface	Their approach regularises the between class and within class matrices with intercluster and intracluster matrices. The cluster-based scatter matrices are estimated from unsupervised clustering.
Coupled marginal discriminant mappings.	Zhang et al. [98]	AR	Data points in the original low and high-resolution features extend into a unified space, where classification is carried out

Fig. 4 Changes in expression adapted from the JAFFE dataset

the scale-space structure for which the information rate is known. Entropy is considered higher order statistical feature, hence provides the least effect on uncertain alterations in the face. The corresponding features are applied to the support vector machine for classification. Experiments were carried out for different plastic surgery conditions on the plastic surgery dataset. They reported that on blepharoplasty, a recognition rate of 88% was achieved, while on brow lift and liposhaving, a recognition rate of 87% and 86% were achieved respectively. In [17], the authors utilized an edge based Gabor feature representation technique for recognizingsurgically changed faces. The edge information that is dependent on the shape of important components of the face is used to address the difference in texture caused by the plastic surgery. Also, in [2], the authors joined the duo of GIST global descriptor and local binary patterns that are feature-based and texture based respectively. Firstly, the local binary pattern was appliedinimportant regions of the face image rather than the entire face image. The concept is based on the idea that only those LBP with necessary information, such as corner and edge will be useful for recognizing faces after plastic surgery. The features extracted by the GIST descriptor is used to get a primary and subordinate level description of the perceptual dimension. Their experiments on the plastic surgery dataset reported 81% recognition accuracy. Table 4 further shows a summary of recent approaches to the issue of plastic surgery.

2.5 Ageing

The problem of ageing in the face recognition system is seen as a type of within-class appearance variation in the faces of human and occurs where a much time difference exists between the targeted face image of the same individual [59]. Ageing variation can have a significant effect on the overall facial structure of individuals as shown in Fig. 6. Also, unique facial characteristics can also change due to changes in the age. Hence, face recognition

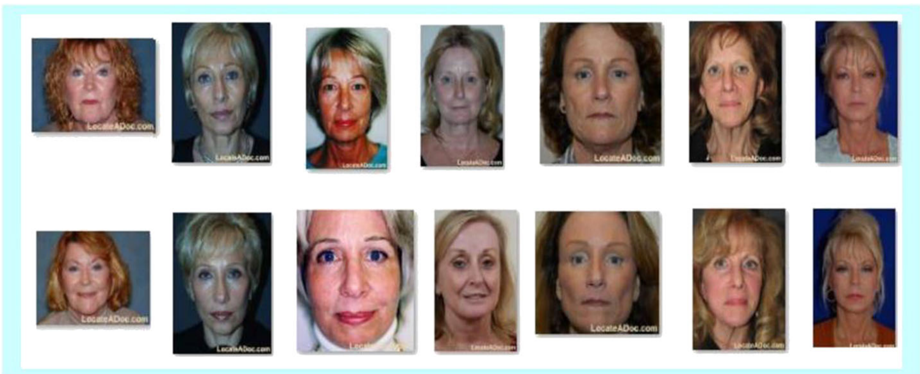

Fig. 5 Different images of the same individual adapted from the plastic surgery dataset

techniques tend to fail in such situations. In literature, it is observed that low performance in ageing occurs greatly when there is a massive gap increase in age between the target image and the query image [51]. Datasets used to evaluate face recognition techniques on the issue of ageing include FG-net and MORPH. These datasets contain a few images and contain other uncontrolled issues such as illumination and pose.

In [104], the authors proposed an identity inference model that is based on age-subspace learning from appearance-age labels. Firstly, the human identity and ageing variables are modeled using probabilistic linear discriminant analysis. Thereafter, the ageing subspace is trained independently with the appearance-age labels, and the identity subspace is then determined with the expectation-maximization algorithm. They observed that the trained ageing subspace is insensitive to the face images utilized. Experiments were carried out on the FGNET, and MORPH dataset achieved a recognition rate of 70.1% and 81.2% respectively. In [92], motivated by the fact that age difference is a nonlinear but smooth transform and the strength of autoencoder network to learn hidden patterns from inputs, authors present a neural network labeled coupled auto-encoder networks for face recognition. Experiments carried out on FGNET dataset showed a recognition rate of 86.5%. Table 5 further shows a summary of recent approaches to the issue of ageing.

2.6 Occlusion

Occlusion can be defined as the hindrance or blockage of a particular section of an image or object. Hence, occlusion of the face image can be described as when part of the face is covered or obstructed either intentionally or unintentionally, thus resulting in a reduction of the performance of the face recognition system [66, 100]. Crimes have always been associated with activities such as terrorist attack, theft at the ATMs and burglary. These activities are mostly conducted by individuals who have their faces occluded such that even installed cameras do not help in identifying them. The blockage of the face image with scarves, sunglasses or hats as shown in Fig. 7. can hinder the performance of the face recognition system as quite substantial information of the face image will be lost. Also, occlusion results to change in the appearance of the face image [25]. Hence, it is essential face recognition systems need to be designed in a way that they are robust to the issue of occlusion.

In [102], the authors proposed a pixel-level occlusion detection based on sparse representation. The sparse representation classification is applied to represent the query image and obtain the residuals of each class, where each pixel is estimated as either occluded or not in each class's residual. A dilation to remove the isolated occlusion estimation pixel is performed, since occlusion us contiguous. They reported a recognition rate of 87.6% on the AR dataset. In [24], the authors proposed an efficient locally-constrained occlusion coding technique that improves the sparse error correction with Markov random field algorithm. Their approach removes

Fig. 6 Age difference of the same individual adapted from MORPH dataset

Fig. 7 Samples of occluded images adapted from AR dataset

occluded sections by locally constrained coding and reduces running time for recognition. Experiments done on the Yale B face dataset reported a recognition rate of 81.9%. Authors in [55], presented a double occlusion scenario where occluded face images appear both in the training and test data. A fuzzy max-pooling approach based on the convolutional neural network was proposed, where unreliable local features are suppressed from occluded regions. An average pooling is further used to enhance the robustness by automatically weighting on each subclass. Experiments carried out on the AR dataset reported a recognition rate of 92.5%. Table 6 further shows a summary of recent approaches to the issue of occlusion.

2.7 Low resolution

The issue of low resolution in face recognition systems occur when the test face image has been degraded drastically. This results inlosingimportant information on the face image across different individuals. This issue affects face recognition systems especially in applications such as surveillance [49]. Hence, it remains an issue in face recognition systems as compared with images in high resolution. In [16], authors investigated low-resolution face recognition with single sample per individual, and present a cluster-based regularized simultaneous discriminant analysis technique. Their approach regularises the between class and within class matrices with intercluster and intracluster matrices. The cluster-based scatter matrices are estimated from unsupervised clustering. Experiments were carried out on the SCface dataset, and they reported a recognition rate of 82.36%. Authors in [98], proposed a useful coupled distance metric learning algorithm labeled coupled marginal discriminant mappings. Their approach makes the data points in the original low and high-resolution features extend into a unified space, where classification is carried out. Experiments were done on the AR face dataset, and a recognition rate of 84.75% was reported. Table 7 further shows a summary of recent approaches to the issue of low resolution.

3 Face recognition techniques

Researchers have proposed many face recognition techniques because the subject area has been a fast-growing and tasking research topic in the field of image processing and computer vision [15]. These techniques have shown satisfactory performance, especially in well-controlled conditions, while there is a drop in performance in real-life conditions. In this section, review on different techniques within the stages of FRS as shown in Fig. 1 are discussed, highlighting their pros and cons where necessary.

3.1 Feature extraction techniques

The feature extraction stage of the FRS involves the simplification of the amount of resources that describes a large set of data. Feature extraction is used to minimize the original face dataset by getting some properties that can be used to classify and get patterns that are present in the input facial images. The different methods of features that can be extracted from the facial images have been presented in the following sub-sections.

3.2 Principal component analysis

Principal Component Analysis (PCA) is a known face recognition technique also called the Eigenface or Karhunen-Loeve expansion [71]. Sirovich and Kirby in 1987 were the first to adequately represent pictures of the human face using PCA after stating that face images can be modified through a minimum aggregation of weights for every face and a standard face image. After that, Turk and Pentland in 1991 proposed the Eigenfaces method for face recognition [88]. The PCA follows a given principle: given a set of training images each having a uniform pixel size, the standard features are extracted from the face image to be left with unique features by subtracting the average face vector from the face vector. Thereafter, the eigenvectors will be calculated from a covariance with reduced dimensionality which is arranged in a way that corresponds to the Eigenvalues, i.e., the larger eigenvalue indicates that the associated eigenvector gets more of the data variance. Figure 8. shows an example of eigenfaces constructed.

3.3 Linear discriminant analysis

Linear Discriminant Analysis (LDA) is also referred to as the Fisher's Discriminant Analysis, and it is also one of the most used techniques in the area of face recognition. Unlike the PCA

Fig. 8 Example of constructed Eigen faces [71]

that builds a subspace to represent the face, the LDA constructs a subspace to differentiate the faces of different people [15]. LDA enables evaluation of the vital information in various parts of the face for recognizing the human face. The primary purpose of LDA is to classify the face image into one or more groups based on features that best describe the face image. Its principle includes: - i) collecting of images and classifying into classes, ii) calculate the vector of the classes where its goal is to maximize the between-class matrix and minimize the within-class matrix, iii) calculate the Eigenvector and iv) obtain the Fisher faces, as shown in Fig. 9. The majority of linear discriminant analysis face recognition system approaches suffer from the fact that their optimality criteria are not directly related to the classification ability of the obtained feature representation [50].

3.4 Independent component analysis

Similar to the PCA, Independent Component Analysis (ICA) is also a known and used subspace method that projects data from a high dimensional space to its lower form. ICA is a feature extraction method that has been considered as a generalization of the PCA that is mainly used to solve problems related to signal processing [15]. ICA is seen as a method that can be employed in the face recognition task where vital information can be contained in the high order relationship amongst pixels [5]. PCA considers images as variables that are random with a Gaussian distribution and minimized second-order statistics. Hence, for a non-Gaussian distribution, large variances will not be matched to PCA basis vectors. The ICA method

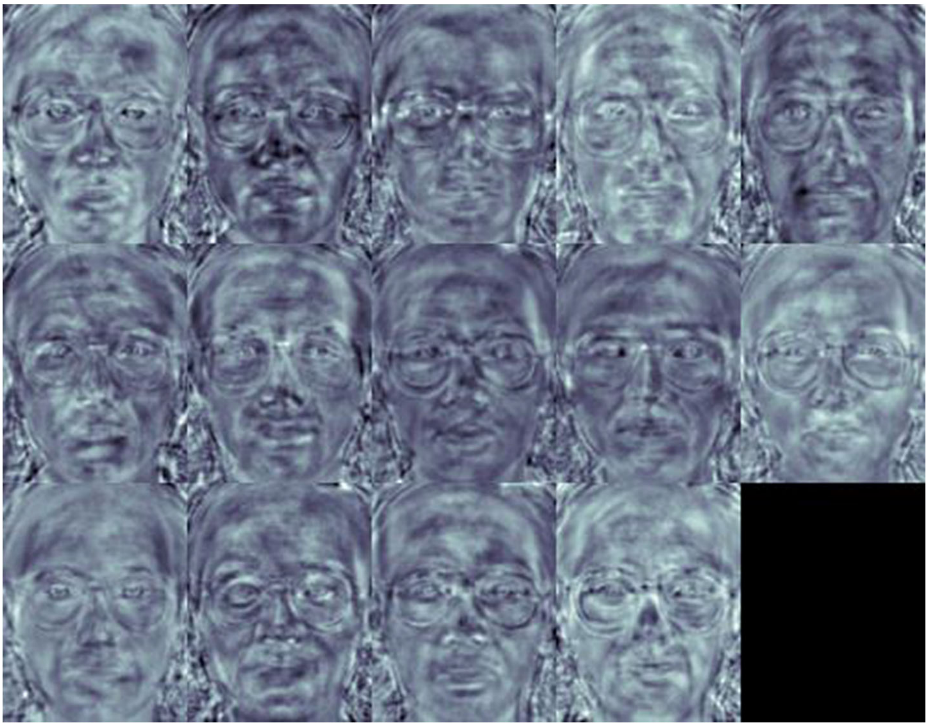

Fig. 9 Example of fisher faces [91]

reduces both second-order and higher-order dependencies in the input data and tries to find a basis in which the data are statistically dependent [18, 35].

3.5 Local binary patterns

The local binary pattern (LBP) operator is described as an image operator that can transform an image into an array of integer labels that further describes the small-scale appearance of the image. The LBP operator was initially designed for the description of texture where the operator gives a label to every pixel of an image by creating binary images from the greyscale image of the three by three neighborhood of each pixel with the middle pixel value and considering the outcome as a binary digit [36]. The LBP labels a point as the center point and then calculates the difference between said points and the points around. If it happens,the difference is greater than zero a one is assigned otherwise it remains zero. Fig. 10 shows the illustration of a generalized LBP operator. Furthermore, it can generate a texture descriptor using the histogram of labels.

3.6 Dynamic link architecture

The dynamic link architecture (DLA) is a neural information processing concept that was first proposed in 1981 as an attempt to solve issues encountered by the artificial neural network (ANN) such as the expression of syntactical relationships [44]. The basic idea of the DLA is the use of synaptic plasticity already present on the time scale of information processing and not only for the acquisition of memory that allows it to immediately group sets of neurons into higher symbolic units [45]. The ability of the DLA is best used in issues of face recognition such as facial expression and invariant object recognition. A DLA that is based on multiscale morphological dilation-erosion was presented for FR to yield a feature vector to verify faces of individuals from a given test set. It showed from their experiment that the presented method performed better as compared to the dynamic link matching based on Gabor wavelets [85].

3.7 Elastic bunch graph matching

The elastic bunch graph matching (EBGM) technique compares images where the algorithms firstly identify landmark locations on images that have look-alike features of the face such as the nose, mouth, and eyes; these features are described using the Gabor wavelet convolution which is known as a Gabor jet when all the values are at a single point. Furthermore, the face graph is used to depict each image where the nodes are placed at the landmark locations with each containing a Gabor jet extracted from that location [8]. There is difficulty in identifying faces from an extensive dataset of images due to image variation regarding size, position, pose and expression. The EBGM is regarded as one of the most successful FR techniques as it has been applied to some face recognition tasks. However, it is required that the modern face

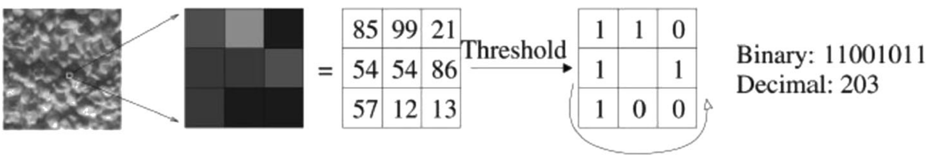

Fig. 10 A basic local binary pattern operator [36]

recognition system be automated-based without any intervention from the side of human. Hence, the drawback of the EBGM lies with the fact that landmark selection of the face image is made manually at the initial stage of the recognition process [46]. Figure 11 shows a sample of an elastic bunch graph matching.

3.8 Geometric feature matching

Geometric feature matching (GFM) methods were one of the earliest ways to recognize individuals using the face. The face image can be recognized even though the details of its major features are not resolved. The remaining information is purely geometrical and represents what is left at a very coarse resolution (i.e., loss of image pixels). However, the idea is to get the informative features like chin and other important facial parts. This method uses a set of training face images to find the eye position in the image and the correlation coefficients is calculated. It further compares this with the test image and searches the maximum values [11].

3.9 Feature classification techniques

The feature classification stage of the FRS involves the process where the acquired facial feature data are categorized into different and given number of classes depending on a particular task. The different methods of feature classification techniques are presented in the following sub-sections:

3.10 Artificial neural network

Artificial Neural Network (ANN) is referred to as a system of interconnected artificial neurons that can share messages amongst one another and also learn from experience. They are inspired by the biological neural system where the computational unit of the brain is a set of interconnected neurons with synapses [77]. The ANN is seen as an efficient and robust classification algorithm that is used for predicting both known and unknown data. The interconnected neurons consist of numeric weights that are adjusted during the training stage in such a way that a well-trained network will efficiently respond when given a recognition task [47]. It also includes multiple layers of feature detecting neurons, where every layer has various neurons that respond to combinations of inputs from the previous layer [94]. These layers are built in a way that the first layer detects a set of given inputs, and subsequent layers detect “patterns of patterns” of the previous layers.

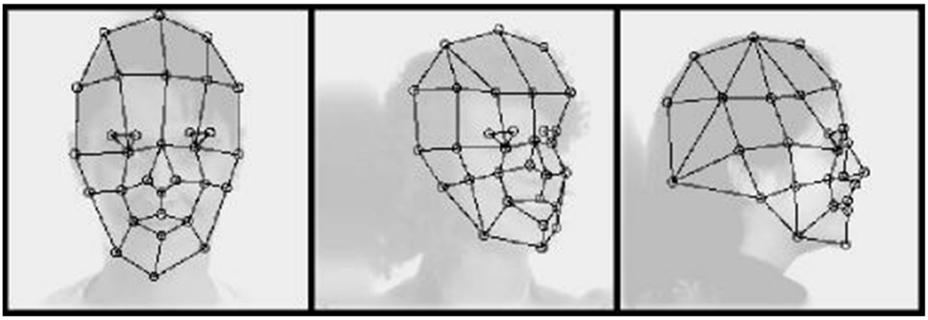

Fig. 11 Example of an elastic bunch graph mathcing [30]

3.11 Convolutional neural network

The Convolutional Neural Network (CNN) is seen to be inspired by the biological evidence found in the visual cortex of the mammalian brain. In the visual cortex, there is the presence of small regions of cells that are sensitive to particular regions of the visual field. An experiment by Hubel and Wisel further displayed that some particular neuronal cells present in the brain triggered only when they notice the edges of a certain orientation. For instance, some neurons trigger to the vertical edges while others trigger to the diagonal or horizontal edges, where the neurons are placed together to execute visual perception. This concept is the basis of the CNN [99]. The CNN was first introduced in the 1980s and the 90s but was neglected due to its complexity on real-world applications. However, it was interestingly revived lately and has since outperformed most computer vision techniques while still growing at a fast pace [32]. The CNN architecture consists of various layers such as the convolutional layer, pooling layer, non-linear layers and the fully connected layers. Each layer has its own impact that is responsible for the performance of the CNN network for various recognition tasks [76]. The convolution layer and the pooling layer act as a neural network that extracts features while the fully connected layers act as the classification neural network that functions based on the image features and comes up with an output [43]. Figure 12 shows a generic architecture of the CNN approach to object recognition. The CNN is an approach that attempts to solve the shortcomings of the NN. These shortcomings include the fact that when the input dimension is high, as regards to the number of connections and images, this will make the number of free parameters high, as there will be a connection between each hidden unit and the input layer. Hence, the number of training samples might be small compared to the pattern dimension, i.e. the NN would have high complexity and could potentially lead to over-fitting of data. Thus, the CNN automatically learns local feature extractors and implement the weight sharing principle enabling a reduction in the number of free parameters thereby increasing the performance capacity as compared to the NN architectures [43]. The CNN is regarded as a class of feed-forward ANN that has shown great achievements in the area of image recognition [32].

3.12 Random Forest

The Random Forest (RF) machine learning technique is one that can be used for both classification and regression tasks. It is a supervised classification technique where the

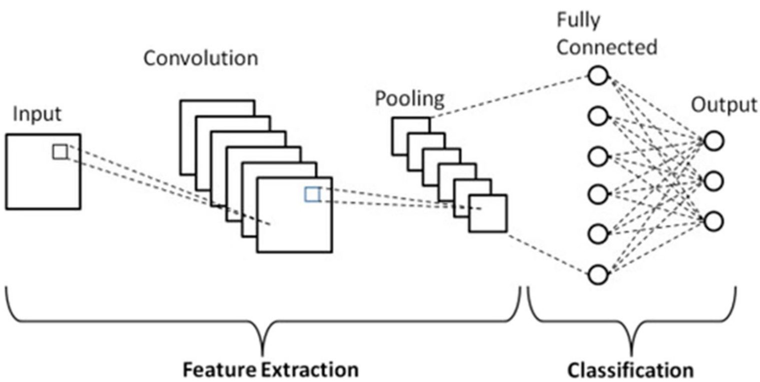

Fig. 12 A generic architecture of the convolutional neural network

approach is created like a forest with some trees as the name implies. The machine learning technique was proposed by Leo Breiman [10], and it is an amalgamation of tree predictors in a way that individual tree is dependent on the values of a random vector tested separately with similar distribution for all trees in the forest. The ability to have random inputs and random features makes RF technique robust, thus capable of dealing with large feature spaces. RF has been used for various computer vision tasks such as object detection, handwritten signature and face recognition due to its low computation complexity, accuracy, its ability to be easily implemented and capability to handle training datasets that are large. RF is a collective classifier that includes some decision trees. An individual tree is made recursively by allocating a binary test to the individual non-leaf node according to the training samples. For classification purposes, the results of the decision tree are combined to select the highest nominated class. The RF classifier could be defined as:

$$H(p) = arg \quad max_q \textstyle\sum_{i=1}^{k} I(hi(p)) = Q$$

Where H (p) is the is the eventually joined classifier, k is the number of decision trees, *hi(p)* depicts a decision tree, Q represents the class label, and I(hi(p) = Q) depicts that p belongs to class Q.

3.13 Support Vector Machine

Support Vector Machine (SVM) has been widely studied over the years and applied for classification and regression tasks where it has shown satisfactory results in detection of faces. The support vector machine (SVM) is a learning method seen to be effective for most pattern recognition tasks due to its ability to perform well without the need of adding another knowledge [73]. If given a set of points that belong to two classes, the SVM locates the hyper plane that can separate the largest possible fractions of points of the same class on the same side, maximizing the distance from either class to the hyper plane. The SVM is also seen as a method to train polynomial neural networks functions classifiers [48]. The principle of the Structure Risk Minimization (SRM) that states the best generalization capabilities are achieved by minimizing the boundary of the generalization error is what these learning techniques are based.. The decision surface is a result of the weighted combination of elements of the training set called the support vectors and portrays the margin amongst the two classes. Assuming the input to an SVM algorithm is defined (pi, qi of labeled training data, in which pi represents the data, and qi equals −1 or 1 and represents the label. Hence, the output of the algorithm depicts a set of support vectors vi; class labels qi, coefficient weights αi and a constant k. Hence, the linear decision surface is thus:

$$w.z + k = 0$$

In which case

$$w = \sum_{i=1}^{Ns} \alpha iviqi$$

Support vector machine can further be extended to a non-linear decision surface by using a kernel C that must satisfy the condition of Mercers. Hence the non-linear decision surface is defined thus:

$$\sum_{i=1}^{Ns} \alpha iviqi(vi, z) + k = 0$$

3.14 Decision table

The Decision Table (DT) is a technique seen as a predictivemodeling tool that carries out the classification task where it does not take into consideration that the attributes are independent. Decision table presents a model that depicts correlations amongst pairs of attributes in an organized hierarchy. Decisions are made by this approach in such a way that the attributes are calculated within the entire level of the tree rather than on a specific sub-tree. Results are further shown in an organized table instead of a tree-like form. Zhou et al. [105] showed how a DT approach for spatial continuity of occlusion couldbe linked into the computation of a sparse representation of the test image concerning the training images. Their experiment indicates that their algorithm further discovers not useful regions and removes them from the sparse representation, thus tolerating fractions and different types of facial occlusion.

3.15 Naïve Bayes

The Naïve Bayes (NB) is a technique based on the probabilities of conditions that use the Bayes' theorem. The Bayes theorem finds the probability by getting the number of frequency values and combinations of values in the dataset. It gets the probability of an act happening to assume the probability of another act is known [38]. Let Y be a training set of multiple elements and corresponding class labels, where each element depicts an n-dimensional attribute vector, V = (V_1 V_2 V_3….V_n), representing n measurement made on the multiple elements from n attributes, respectively, B_1,B_2,B_3…..B_n. Assuming there are n classes, I_1,I_2,I_3…..In. Given a tuple, Z, the classifier will envisage that Z is appropriate to be in a class that has the highest posterior probability, as regards Z. Thus, the NB envisages that tuple Z is of the class I1 only if P (Ii Z) > P (Ij Z) for 1 < j < n, J is not equal to i. The class I_1 for which P (Ii Z) is maximized is known as the maximum posterior hypothesis [26]. Hence by Bayes theorem:

$$P(Ii/Z) = \frac{P(Ii/Z)P(Ii)}{P(Z)}$$

4 Face recognition benchmark datasets

In this section, the major face recognition face datasets that have been developed are discussed. These face datasets include one or more constraints that depicts real-life scenarios which researchers use to evaluate their various face recognition techniques or approaches. A summary

Table 8 Face datasets and conditions in uncontrolled environment

Face Dataset	Scenarios in Uncontrolled Environment						
	Illumination	Pose	Expression	Plastic Surgery	Ageing	Occlusion	Low resolution
AR Face	✓		✓			✓	✓
Yale B	✓	✓					
Life in the Wild	✓	✓	✓			✓	✓
FERET							✓
ORL		✓	✓				
FG-net	✓	✓			✓		
Plastic Surgery	✓			✓			
MS-Celeb-1 M	✓			✓		✓	

Table 9 Result performance from different research works considering specific challenges on the AR face dataset

Research work	Technique	Challenges	Recognition performance
Zhou et al. [107]	Sparse-representation based system that controls dense pixel correspondences between training and testing facial samples.	• Facial Expression • Natural variations of misalignment.	98.7%
Deng et al. [19]	Superposed linear representation classifier (SLRC) to cast the recognition issue by representing the test image in terms of a superposition of the class centroids and the shared intra-class differences.	• Occlusion • Illumination • Expression	97.5%
Ding et al. [21]	Algorithm based on Kernel principal component analysis network (KPCANet) that uses a weighted voting scheme to extract single image features and identify the probes.	• Single sample per person (SSPP) with illumination, and expression.	98.7%, 86.0% respectively.
Jin et al. [40]	A collaborative representation based FR method that balances data locality and collaborative representation is presented. The approach combines a locality adaptor term into the robust collaborative representation based framework, thus leading to a novel unified objective function.	• Occlusion with scarf	78.8%
Tan et al. [84]	A Kernelized locality-sensitive group sparsity representation is presented. The approach considers both the grouped structure information of the training dictionary and the data locality in the kernel space. As a result, the structure and nonlinear information embedded in the training and test data can be better utilized, and more discriminative sparse representation can be obtained.	• Illumination • Expression, under varying PCA dimension reduction at 200.	91.41%

showing the constraints in the various face datasets is presented in Table 8. Furthermore, some performance results on each of the datasets are presented in Tables 9, 10, 11, 12, 13, 14 and 15.

4.1 AR face dataset

The Aleix Martinez (AR) face dataset is a collection of over 4000 colored face images of 126 individuals (56 women and 70 men) obtained by the Computer Vision Center in Barcelona, Spain in 1998. During the collection of these images, the recording and the imaging conditions (camera distance, camera parameters, and illumination settings) were carefully and strictly selected to ensure the settings are the same for different subjects [60]. These images were recorded twice, at a two weeks interval with a pixel size of 768 × 576. During each session, 13 different conditions with varying facial expression, occlusion and illumination were captured. Table 9 shows performance from different research works considering specific challenges on the AR face dataset.

4.2 Yale face dataset B

The Yale face dataset was constructed by the Yale Center for Computation Vision and Control and consists of a collection of 165 images of 15 individuals with each person providing 11

Table 10 Result performance from different research works considering specific challenges on the Yale face dataset B

Research work	Technique	Challenges	Recognition performance
Ding et al. [21]	Algorithm based on Kernel principal component analysis network (KPCANet) that uses a weighted voting scheme to extract single image features and identify the probes.	• Single sample per person (SSPP) with illumination, and 80% facial occlusion.	96.6%, 85% respectively.
Tan et al. [84]	A Kernelized locality-sensitive group sparsity representation is presented. The approach considers both the grouped structure information of the training dictionary and the data locality in the kernel space. As a result, the structure and nonlinear information embedded in the training and test data can be better utilized, and more discriminative sparse representation can be obtained.	• Illumination • Expression, under varying PCA dimension reduction at 54.	94.48%
Cheng et al. [14]	A face recognition model that consists of a two-layer deep convolutional neural network for extraction of features, and Sparse representation classification (SRC) for classification.	• Expression, under varying PCA dimension reduction at 504.	99.17%
Rakshit et al. [97]	Multiple representations of a face image using different local descriptors derived from LBP and Local Graph Structure (LGS). Then max, average and L2 pooling operations are applied on each representation of face image to scale down the features. Lastly, two correlation based, and KNN are used to produce similar proximities for grouping a probe user with rank identity.	• Expression	93.6%

different images. These images are made up of a variety of conditions, such as demonstration of variations in lightning conditions (right-light, center-light and left-light), facial expression (normal, sad, happy, surprised, sleepy and wink) and with or without glasses [53]. Minimal changes in the facial expression and head position are visible because all the 64 images of a face for a particular pose was captured within a time frame of about 2 s. The Yale face dataset has been divided into four different subsets concerning the angle between the light source and the camera axis (120, 250, 500, 770). The hand-labeled locations of both the eyes and the center of the mouth are distributed along with the dataset. Table 10 shows performance from different research works considering specific challenges on the Yale face dataset B.

Due to the need for systematic testing of face recognition methods under large pose and illumination variations, the Yale face dataset B was collected. This dataset contains subjects that were imaged inside a geodesic dome using 64 computer-controlled xenon strobes. These images are from 10 different individuals under 64 lighting conditions and nine different poses (five poses at 120, three poses at 240 from the camera axis and one frontal pose) (110). Minimal changes in the facial expression and head position are visible because all the 64 images of a face for a particular pose was captured within a time frame of about 2 s. The Yale face dataset has been divided into four different subsetsconcerning the angle between the light

Table 11 Result performance from different research works considering specific challenges on the Labelled faces in the wild face dataset

Research work	Technique	Challenges	Recognition performance
Zhou et al. [107]	Sparse-representation based system that controls dense pixel correspondences between training and testing facial samples.	• Facial Expression • Pose variation	95.2%
Tan et al. [84]	A Kernelized locality-sensitive group sparsity representation is presented. The approach considers both the grouped structure information of the training dictionary and the data locality in the kernel space. As a result, the structure and nonlinear information embedded in the training and test data can be better utilized, and more discriminative sparse representation can be obtained.	• Illumination • Expression, under varying PCA dimension reduction at 300.	70.01%
Zhang et al. [101]	A customized weighted discriminative loss to seek a customized constraint for correcting the large error caused by imbalanced distributions of correctly classified and misclassified features.	• Illumination • Pose variation	94.28%
Zhao et al. [103]	Approach focuses on the aspect of discriminability i.e. the inter class distinctiveness that is usually neglected. Furthermore, it explicitly distances identities by penalizing the angle between an identity and its nearest neighbor, resulting in discriminative face representations.	• Pose, and expression taking into consideration scenario where facial images of the same identity is expected to be closer in the representation space.	99.32%

Table 12 Result performance from different research works considering specific challenges on the FERET face dataset

Research work	Technique	Challenges	Recognition performance
Deng et al. [19]	Superposed linear representation classifier (SLRC) to cast the recognition issue by representing the test image in terms of a superposition of the class centroids and the shared intra-class differences.	• Single training sample per person.	96.3%
Ding et al. [21]	Algorithm based on Kernel principal component analysis network (KPCANet) that uses a weighted voting scheme to extract single image features and identify the probes.	• Single sample per person (SSPP) with illumination.	90.3%
Zeng et al. [97]	A sparse representation-based classification approach that implements an anti-noise sparse representation using both l_1 and l_2 to exploit good performance.	• Facial expression with 10% noise, and 4 training samples.	75.5%

Table 13 Result performance from different research works on the plastic surgery face dataset

Research work	Technique	Challenges	Recognition performance
Sabharwal et al. [79]	A region based score level fusion method for local facial features is presented to equalize former and latter surgery face images.	Plastic surgery	87.0%
Rakshit et al. [75]	Multiple representations of a face image using different local descriptors derived from LBP and Local Graph Structure (LGS). Then max, average and L2 pooling operations are applied on each representation of face image to scale down the features. Lastly, two correlation based, and KNN are used to produce similar proximities for grouping a probe user with rank identity.	Plastic surgery	89.6%
Suri et al. [83]	The proposed algorithm combines off-the-shelf supervised classifier and a generic, task independent network which encodes information related to basic visual cues such as color, shape, and texture.	Plastic surgery	92.1%

source and the camera axis (120, 250, 500, 770). The hand-labeled locations of both the eyes and the center of the mouth are distributed along with the dataset.

4.3 Labelled faces in the wild

The Labelled Faces in the Wild (LFW) dataset is a collection of 13,233 images of faces collected from the web and labeled with each person's name. The dataset consists of 5749 different individuals with 1680 of the individuals having at least two or more distinct images in the dataset [37]. The remaining 4069 contains just a single image in the dataset. Most of the images in the LFW dataset are colored and available in 250 × 250 pixels JPEG with only a few grayscale images. The LFW dataset is divided into two; one for algorithm development and

Table 14 Result performance from different research works on the IJB-A face dataset

Research work	Technique	Challenges	Recognition performance
Qian et al. [74]	A face normalization model (FNM) is presented for unsupervised face normalization in condition of unconstrained environment. The FNM uses a face expert network to produce face identity. This decomposes the task of generator to employ high-level sematic feature instead of image. A pixel-wise loss is further introduced by a new way of stabilizing training optimization and high quality result.	• Illumination with facial expression.	96.0%
Tan et al. [84]	A Kernelized locality-sensitive group sparsity representation is presented. The approach considers both the grouped structure information of the training dictionary and the data locality in the kernel space. As a result, the structure and nonlinear information embedded in the training and test data can be better utilized, and more discriminative sparse representation can be obtained.	• Illumination • Expression, under varying PCA dimension reduction at 300.	32.68%
Masi et al. [62]	Deep convolutional neural network is utilized to learn discriminative representations.	• Pose variation	92.8%

Table 15 Result performance from different research works on the CMU Multi-PIE face dataset

Research work	Technique	Challenges	Recognition performance
Qian et al. [74]	A face normalization model (FNM) is presented for unsupervised face normalization in condition of unconstrained environment. The FNM uses a face expert network to produce face identity. This decomposes the task of generator to employ high-level sematic feature instead of image. A pixel-wise loss is further introduced by a new way of stabilizing training optimization and high quality result.	• Illumination and Pose variation at $\pm$60 degrees.	93.7%
Jin et al. [40]	A collaborative representation based FR method that balances data locality and collaborative representation is presented. The approach combines a locality adaptor term into the robust collaborative representation based framework, thus leading to a novel unified objective function.	• Illumination • Expression	96.9%
Petpairote et al. [72]	Utilizing triangle transformation to transform a pose face to a frontal face and an expression face to opened-eyes and closed-mouth face.	• Pose • Expression	92.4%
Liang et al. [13]	A multi-scale parallel convolutional neural network architecture to extract deep robust facial features with high discriminative ability.	• Illumination • Pose variation • Expression	94.5%

the other for performance reporting. Table 11 shows performance from different research works considering specific challenges on the Labelled in the wild face dataset.

4.4 Facial recognition technology

The Facial Recognition Technology (FERET) dataset is a collection of 24 facial image categories, obtained from the George Mason University and the US Army Research Laboratory facility as part of the FERET program which was sponsored by the US Department of Defence Counterdrug Technology and Development Program [53]. FERET has been used extensively in past works, for example, FERET and facial recognition vest test (FRVT) 2000 datasets has been exclusively evaluated and available for use by both commercial facial recognition systems and research algorithms. The list of images in FERET dataset used for training, gallery, and probing is distributed in line with the dataset, therefore ensuring that adequate comparison of proposed recognition algorithms can be extensively carried out. Records obtained show that the FERET dataset has been distributed to over 460 research groups. The facial images in the FERET dataset were recorded between August 1993 and July 1996 over 15 sessions. A 35 mm camera was used for the image recording before subsequently digitalizing the images and then converted to 8-bit gray-scale images of 256 X 384 pixel in size. FERET dataset contains little variations between recording sessions as a result of having to reassemble the recording equipment for each session. The ground-truth information together with the date the recording was carried out and information that indicates if the subject is wearing glasses is supplied for each image in the dataset. Additionally, the locations of the right and left eye and mouth center that have been manually determined is available for 3816 images. The adjustment has been made on the new version of the FERET dataset, as NIST has made a higher resolution color images (512 X 768) of the initial gray-scale images. Table 12 shows performance from different research works considering specific challenges on the FERET face dataset.

4.5 Plastic surgery face dataset

The plastic surgery face dataset is a collection of faces that depicts a real-life scenario of the surgery issue to face recognition system. The facial dataset contains 1800 before and after surgery face images that belong to 900 different individuals. 519 of these individuals represents the situation of local surgeries, i.e. changes in some part of the facial image and 381 individuals representing cases of global surgery, i.e. total change of the face image [2]. The dataset contains different types of plastic surgery, i.e. Rhytidectomy (face lift), Rhinoplasty (surgery of the nose) and Blepharoplasty (surgery of the eyelid). Table 13 shows performance from different research works considering specific challenges on the plastic surgery face dataset.

4.6 IARPA Janus benchmark a (IJB-A)

IJB-A is one of the most challenging unconstrained face recognition benchmark dataset with uncontrolled pose variations. IJB-A contains both images and video frames from 500 subjects with 5397 images and 2042 videos that are split into 20,412 frames, captured from in-the-wild environment to avoid the near frontal bias, along with protocols for evaluation of both verification (1:1 comparison) and identification (1:N search) tasks. Table 14 shows result performance from different research works on the Plastic surgery face dataset. Table 14 shows performance from different research works considering specific challenges on the IJB-A face dataset.

4.7 CMU multi-PIE

The CMU Multi-PIE dace dataset contains up to 750,000 face images of 337 subjects that have been recorded in a number of four sessions over a duration five months. Individuals were imaged under 15 different pose and 19 illumination conditions while displaying varying facial expressions. Also, high resolution frontal face images were acquired [31]. Table 15 shows performance from different research works considering specific challenges on the CMU Multi-PIE face dataset.

5 Future research directions

In the preceding sections, we have described an up to date status of the research on face recognition systems. The various issues of the face recognition system also termed face recognition in the wild have been discussed i.e. illumination, pose, expression, plastic surgery, ageing, occlusion and low resolution. It is observed that out of the mentioned constraints the issue on pose and expression are the most researched about in literature while other issues such as occlusion, illumination, and low resolution are less researched. This is because not enough face datasets depict these issues of real-life scenarios. Hence, facial datasets that can address these issues should be developed to enable researchers come up with algorithms that can be robust to these constraints. Also, following the critical review of this paper, the following can be considered as future research directions:

Pre-processing approaches: most face recognition approaches concentrate more on the feature extraction and classification stages, and as result neglect specific pre-processing steps that can enhance face recognition systems. For example, face alignment has in recent times gained attention due to its comprehensive application in automatic face analysis. Though, it has shown to be extremely challenging in unconstrained environments. Also, face image enhancement is another pre-processing step that can help improve the performance of face recognition techniques [68]. A major setback with the process is the ability to extract effective features from the enhanced face images that can be effective for classification [67].

Deep neural networks: Deep neural networks have shown to be effective techniques for computer vision tasks and most recently for face recognition tasks. Notwithstanding, the attempts of previous authors to solve facial constraints during face recognition, there still exist the possibility of improving individual layers of the CNN that can make the network perform optimally. The pooling layer of the CNN architecture is an essential layer because it can reduce the size of the feature map that portrays the filtered image in the convolution layer. Also, it enhances the performance of the CNN by transforming invariances such as expansion, rotation, and translation. Hence, it is one such component of the CNN architecture that can be used to address the significant issues of face recognition to achieve optimal performance. The pooling layer consists of both the max and average pooling, therefore, the performance of the CNN can largely depend on the type of pooling method used; which is also a function of the input image [86]. Even though the max and average pooling method perform well on certain datasets, it is unsure which pooling method will outperform the other on different face recognition problems. Also, other approaches can be introduced into the pooling layer on how to select features that can perform efficiently.

Face datasets: The facial datasets are discussed describing the various issues that each contains. Access to the facial datasets is somewhat difficult as much processes is involved, thus making research in the area slow. Also, face datasets should be developed such that the collection of faces can include most of the facial constraints. By doing this, a face recognition algorithm can be evaluated with just a single face dataset that depicts a typical real-life scenario. Lastly, it is seen from the results reported on the different face datasets that continuous research work on improving face recognition algorithms needs to be done.

6 Conclusion

In this paper, we reviewed academic literature on face recognition techniques that have been developed over time. After that, recently published academic literature between 2015 to date were also reviewed to appreciate the level of work that has been carried out in the field of face recognition systems and also presents a state-of-the-art performance. We also presented the various issues of the face recognition system labeled face recognition in the wild or face recognition in an unconstrained environments. These issues were described, and reasons where they continually need to be researched on are stated. Also, the various techniques discussed in the literature have been presented. Furthermore, we have summarised the various techniques approach to develop a robust face recognition system capable of solving different real-life scenarios. However, optimal performance of the various techniques on the different face recognition constraints has not been achieved. The face datasets that contain major constraints of the face recognition system are discussed, and state-of-the-art results on each datasets are presented. We recommend the need for easy access to face recognition datasets while also suggesting that a face dataset can be developed such that most of the major constraints are designed within it. This paper also presents vital areas for future research directions, and finally, the paper has been articulated in such a way to benefit new and existing researchers in this field.

Acknowledgments This work was supported by the Council for Scientific and Industrial Research (CSIR), South Africa.

[ICT: Meraka].

References

1. Abate AF, Nappi M, Riccio D, Sabatino G (2007) 2D and 3D face recognition: a survey. Pattern Recogn Lett 28:1885–1906
2. Ali ASO, Sagayan V, Malik A, Aziz A (2016) Proposed face recognition system after plastic surgery. IET Comput Vis 10:344–350
3. Alkkiomaki O, Kyrki V, Liu Y, Handroos H, and Kalviainen H (2009) Multi-modal force/vision sensor fusion in 6-DOF pose tracking," in Advanced Robotics. ICAR 2009. International conference on 2009,, pp. 1–8.
4. Angadi SA, Kagawade VC (2017) A robust face recognition approach through symbolic modeling of polar FFT features. Pattern Recogn 71:235–248
5. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13:1450–1464
6. Belahcene M, Chouchane A, and Ouamane H (2014) 3D face recognition in presence of expressions by fusion regions of interest," in 2014 22nd Signal Processing and Communications Applications Conference (SIU), pp. 2269–2274.
7. Bhat FA, Wani MA (2016) Elastic bunch graph matching based face recognition under varying lighting, pose, and expression conditions. IAES International Journal of Artificial Intelligence (IJ-AI) 3:177–182
8. Bolme DS (2003) Elastic bunch graph matching. Colorado State University
9. Bowyer KW, Chang K, Flynn P (2006) A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition. Comput Vis Image Underst 101:1–15
10. Breiman L (2001) Random forests. Mach Learn 45:5–32
11. Brunelli R, Poggio T (1993) Face recognition: features versus templates. IEEE Trans Pattern Anal Mach Intell 15:1042–1052
12. Cao X, Shen W, Yu L, Wang Y, Yang J, Zhang Z (2012) Illumination invariant extraction for face recognition using neighboring wavelet coefficients. Pattern Recogn 45:1299–1305
13. Chen L, Liang M, Song W, and Xiao K (2018) A multi-scale parallel convolutional neural network based intelligent human identification using face information. Journal of Information Processing Systems, vol. 14.
14. Cheng EJ, Chou KP, Jin S, Tanveer M, Lin CT, Young KY, Lin WC, Prasad M (2019) Deep sparse representation classifier for facial recognition and detection system. Pattern Recogn Lett 125:71–77
15. Chihaoui M, Elkefi A, Bellil W, Ben Amar C (2016) A survey of 2D face recognition techniques. Computers 5:21
16. Chu Y, Ahmad T, Bebis G, Zhao L (2017) Low-resolution face recognition with single sample per person. Signal Process 141:144–157
17. Chude-Olisah CC, Sulong G, Chude-Okonkwo UA, Hashim SZ (2014) Face recognition via edge-based Gabor feature representation for plastic surgery-altered images. EURASIP Journal on Advances in Signal Processing 2014:102
18. Delac K, Grgic M, Grgic S (2005) Independent comparative study of PCA, ICA, and LDA on the FERET data set. Int J Imaging Syst Technol 15:252–260
19. Deng W, Jiani H, Jun G (2017) Face recognition via collaborative representation: its discriminant nature and superposed representation. IEEE Transaction on pattern analysis and machine intelligence 40:2513–2521
20. Ding C, Tao D (2017) Pose-invariant face recognition with homography-based normalization. Pattern Recogn 66:144–152
21. Ding C, Hu Z, Karmoshi S, Zhu M (2017) A novel two-stage learning pipeline for deep neural networks. Neural processing letters

22. Drira H, Amor BB, Srivastava A, Daoudi M, Slama R (2013) 3D face recognition under expressions, occlusions, and pose variations. IEEE Trans Pattern Anal Mach Intell 35:2270–2283
23. Feng Z-H, Kittler J, Awais M, Huber P, and Wu X-J (2017) Face detection, bounding box aggregation and pose estimation for robust facial landmark localisation in the Wild, arXiv preprint arXiv:1705.02402.
24. Fu Y, Wu X, Wen Y, Xiang Y (2017) Efficient locality-constrained occlusion coding for face recognition. Neurocomputing 260:104–111
25. Gao G, Yang J, Jing X-Y, Shen F, Yang W, Yue D (2017) Learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn 66:129–143
26. Gao C-z, Cheng Q, He P, Susilo W, Li J (2018) Privacy-preserving naive Bayes classifiers secure against the substitution-then-comparison attack. Inf Sci 444:72–88
27. Ghiass RS, Arandjelović O, Bendada A, Maldague X (2014) Infrared face recognition: a comprehensive review of methodologies and datasets. Pattern Recogn 47:2807–2824
28. Goyal SJ, Upadhyay AK, Jadon R, and Goyal R (2018) Real-life facial expression recognition systems: a review," in Smart Computing and Informatics, ed: Springer, pp. 311–331.
29. Guo Y, Zhang L, Hu Y, He X, and Gao J (2016) Ms-celeb-1m: A dataset and benchmark for large-scale face recognition, in European Conference on Computer Vision, pp. 87–102.
30. Hanmandlu M, Gupta D, and Vasikarla S (2013) Face recognition using Elastic bunch graph matching. in Applied Imagery Pattern Recognition Workshop (AIPR): Sensing for Control and Augmentation, 2013 IEEE, pp. 1–7.
31. Heo J, Marios S (2008) Face recognition across pose using view based active appearance models on CMU multi-PIE dataset. In Proceeding of International Conference on Computer Vision Systems, May, pp 527–535
32. Hijazi S, Kumar R, and Rowen C (2015) Using convolutional neural networks for image recognition, ed.
33. Ho C, Morgado P, Persekian A, Vasconcelos N (2019) "PIEs: pose invariant Embeddings," IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach, CA, USA, pp 12369–12378. https://doi.org/10.1109/CVPR.2019.01266
34. Hsu G-SJ, Ambikapathi A, Chung S-L, Shie H-C (2018) Robust cross-pose face recognition using landmark oriented depth warping. J Vis Commun Image Represent 53:273–280
35. Hu H (2008) ICA-based neighborhood preserving analysis for face recognition. Comput Vis Image Underst 112:286–295
36. Huang D, Shan C, Ardabilian M, Wang Y, Chen L (2011) Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Syst Man Cybern Part C Appl Rev 41:765–781
37. Jia S, Lansdall-Welfare T, and Cristianini N (2016) Gender classification by deep learning on millions of weakly labelled images, in Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on, pp. 462–467.
38. Jiang L, Li C, Wang S, Zhang L (2016) Deep feature weighting for naive Bayes and its application to text classification. Eng Appl Artif Intell 52:26–39
39. Jin X, Tan X (2017) Face alignment in-the-wild: a survey. Comput Vis Image Underst 162:1–22
40. Jin T, Liu Z, Yu Z, Min X, Li L (2017) Locality preserving collaborative representation for face recognition. Neural Process Lett 45:967–979
41. Kakadiaris IA, Toderici G, Evangelopoulos G, Passalis G, Chu D, Zhao X, Shah SK, Theoharis T (2017) 3D-2D face recognition with pose and illumination normalization. Comput Vis Image Underst 154:137–151
42. Karamizadeh S, Abdullah SM, Zamani M, Shayan J, and Nooralishahi P (2017) Face recognition via taxonomy of illumination normalization," in Multimedia Forensics and Security, ed: Springer, pp. 139–160.
43. Kim P (2017) Convolutional Neural Network, in MATLAB Deep Learning, ed: Springer, pp. 121–147.
44. Kotropoulos C, Pitas I, Fischer S, and Duc B (1997) Face authentication using morphological dynamic link architecture," in Audio-and Video-based Biometric Person Authentication, pp. 169–176.
45. Lades M, Vorbruggen JC, Buhmann J, Lange J, von der Malsburg C, Wurtz RP, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. IEEE Trans Comput 42:300–311
46. Lahasan BM, Venkat I, Al-Betar MA, Lutfi SL, De Wilde P (2016) Recognizing faces prone to occlusions and common variations using optimal face subgraphs. Appl Math Comput 283:316–332
47. Le QV (2013) Building high-level features using large scale unsupervised learning, in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pp. 8595–8598.
48. Lei G, Li X-h, Zhou J-l, and Gong X-g (2009) Geometric feature based facial expression recognition using multiclass support vector machines," in Granular Computing, 2009, GRC'09. IEEE International Conference on, pp. 318–321.
49. Li L-y, Li D-r (2010) Research on particle swarm optimization in remote sensing image enhancement [J]. Journal of Geomatics Science and Technology 2:012

50. Li M, Yuan B (2005) 2D-LDA: a statistical linear discriminant analysis for image matrix. Pattern Recogn Lett 26:527–532
51. Li Z, Gong D, Li X, Tao D (2016) Aging face recognition: a hierarchical learning model based on local patterns selection. IEEE Trans Image Process 25:2146–2154
52. Li Y, Wang Y, Liu J, Hao W (2018) Expression-insensitive 3D face recognition by the fusion of multiple subject-specific curves. Neurocomputing 275:1295–1307
53. Liao S, Lei Z, Yi D, and Li SZ (2014) A benchmark study of large-scale unconstrained face recognition," in Biometrics (IJCB), 2014 IEEE International Joint Conference on, pp. 1–8.
54. Liu H-D, Yang M, Gao Y, Cui C (2014) Local histogram specification for face recognition under varying lighting conditions. Image Vis Comput 32:335–347
55. Long Y, Zhu F, Shao L, and Han J (2017) Face recognition with a small occluded training set using spatial and statistical pooling. Inf Sci.
56. Lopes AT, de Aguiar E, De Souza AF, Oliveira-Santos T (2017) Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recogn 61: 610–628
57. Luan X, Fang B, Liu L, Yang W, Qian J (2014) Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion. Pattern Recogn 47:495–508
58. Ma X, Song H, Qian X (2015) Robust framework of single-frame face Superresolution across head pose, facial expression, and illumination variations. IEEE Transactions on Human-Machine Systems 45:238–250
59. Manjani I, Sumerkan H, Flynn PJ, and Bowyer KW (2016) Template aging in 3D and 2D face recognition, " in Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, pp. 1–6.
60. Martinez AM (1998) The AR face dataset, CVC technical report, vol. 24.
61. Martins JA, Lam R, Rodrigues J, du Buf J (2018) Expression-invariant face recognition using a biological disparity energy model. Neurocomputing 297:82–93
62. Masi L, Rawls S, Medioni G, and Natarajan P (2016) Pose-aware face recognition in the wild. In Proceedings of the IEEE Conference on computer vision and pattern recognition, pp. 4838–4846.
63. Mi J-X, Liu T (2016) Multi-step linear representation-based classification for face recognition. IET Comput Vis 10:836–841
64. Nappi M, Ricciardi S, Tistarelli M (2016) Deceiving faces: when plastic surgery challenges face recognition. Image Vis Comput 54:71–82
65. Oloyede MO, Hancke GP (2016) Unimodal and multimodal biometric sensing systems: a review. IEEE Access 4:7532–7555
66. Oloyede MO, Hancke GP, and Kapileswar N (2017) Evaluating the effect of occlusion in face recognition systems, In Proceedings of IEEE Africon Conference, pp. 1547–1551.
67. Oloyede MO, Hancke GP, and Myburgh HC (2018) Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network. IEEE Access, pp. 1–11.
68. Oloyede MO, Hancke GP, Myburgh HC, and Onumanyi AJ (2019) A new evaluation function for face image in unconstrained environments using metaheuristic algorithms. Eurasip Journal on Image and Video Processing, pp. 1–18.
69. Ouyang S, Hospedales T, Song Y-Z, Li X, Loy CC, Wang X (2016) A survey on heterogeneous face recognition: sketch, infra-red, 3d and low-resolution. Image Vis Comput 56:28–48
70. Patacchiola M, Cangelosi A (2017) Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. Pattern Recogn 71:132–143
71. Pereira JF, Barreto RM, Cavalcanti GD, and Tsang R (2011) A robust feature extraction algorithm based on class-modular image principal component analysis for face verification, in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 1469–1472.
72. Petpairote C, Madarasmi S, Chamnongthai K (2017) A pose and expression face recognition method using transformation based on single face neutral reference. In Proceedings of IEEE Internationl Conference on Global Wireless Summit, October:123–126
73. Qi Z, Tian Y, Shi Y (2013) Robust twin support vector machine for pattern classification. Pattern Recogn 46:305–316
74. Qian Y, Deng W, and Hu J (2019) Unsupervised face normalization with extreme pose and expressionin the wild , In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9851–9858.
75. Rakshit RD, Kisku DR (2020) Face identification via strategic combination of local features. In Proceedings of Computational Intelligencein Pattern Recognition:207–217

76. Rasti P, Uiboupin T, Escalera S, and Anbarjafari G (2016) Convolutional neural network super resolution for face recognition in surveillance monitoring, in International Conference on Articulated Motion and Deformable Objects, pp. 175–184.
77. Rehman A, Saba T (2014) Neural networks for document image preprocessing: state of the art. Artif Intell Rev 42:253–273
78. Revina IM, Emmanuel WS (2018) Face expression recognition using LDN and dominant gradient local ternary pattern descriptors. Journal of King Saud University-Computer and Information Sciences
79. Sabharwal T, Rashimi G (2019) Human identification after plastic surgery using region based score level fusion of local facial features. Journal of information security and application 48:102373
80. Sable AH, Talbar SN, Dhirbasi HA (2017) Recognition of plastic surgery faces and the surgery types: An approach with entropy based scale invariant features. Journal of King Saud University-Computer and Information Sciences
81. Sariyanidi E, Gunes H, Cavallaro A (2015) Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans Pattern Anal Mach Intell 37:1113–1133
82. Savran A, Sankur B (2017) Non-rigid registration based model-free 3D facial expression recognition. Comput Vis Image Underst 162:146–165
83. Suri S, Sankaran A, Vasta M, Singh R (2018) On matching faces with alterations due to plastic surgery and disguise. In Proceedings of IEEE Conference on Biometrics Theory, Applications and Systems, pp 1–7
84. Tan S, Xi S, Wenato C, Lei Q, Ling S (2017) Robust face recognition with kernalized locality-sensitive group sparsity representation. IEEE Transaction on image processing 26:4661–4668
85. Tefas A, Kotropoulos C, and Pitas I (1998) Variants of dynamic link architecture based on mathematical morphology for frontal face authentication, in Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231), pp. 814–819.
86. Tong Z, Aihara K, and Tanaka G (2016) A hybrid pooling method for convolutional neural networks, in International Conference on Neural Information Processing, pp. 454–461.
87. Tsai H-H, Chang Y-C (2017) Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Comput:1–17
88. Turk MA and Pentland AP (1991) Face recognition using eigenfaces, in Computer Vision and Pattern Recognition. Proceedings CVPR'91., IEEE Computer Society Conference on, 1991, pp. 586–591.
89. Wang K, Chen Z, Wu QJ, Liu C (2017) Illumination and pose variable face recognition via adaptively weighted ULBP_MHOG and WSRC. Signal Process Image Commun 58:175–186
90. Wang J-W, Le NT, Lee J-S, Wang C-C (2017) Illumination compensation for face recognition using adaptive singular value decomposition in the wavelet domain. Inf Sci
91. Xanthopoulos P, Pardalos PM, and Trafalis TB (2013) Linear discriminant analysis, in Robust data mining, ed: Springer, pp. 27–33.
92. Xu C, Liu Q, Ye M (2017) Age invariant face recognition and retrieval by coupled auto-encoder networks. Neurocomputing 222:62–71
93. Yang J, Luo L, Qian J, Tai Y, Zhang F, Xu Y (2017) Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. IEEE Trans Pattern Anal Mach Intell 39:156–171
94. Yang J, Ren P, Zhang D, Chen D, Wen F, Li H, and Hua G (2017) Neural aggregation network for video face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4362–4371.
95. Yu Y-F, Dai D-Q, Ren C-X, Huang K-K (2017) Discriminative multi-layer illumination-robust feature extraction for face recognition. Pattern Recogn 67:201–212
96. Zafeiriou S, Zhang C, Zhang Z (2015) A survey on face detection in the wild: past, present and future. Comput Vis Image Underst 138:1–24
97. Zeng S, Jianping G, Deng L (2017) An antinoise sparse representation method for robust face recognition via joint $l_1$and l_2 regularization. Expert System with Application 82:1–9
98. Zhang P, Ben X, Jiang W, Yan R, Zhang Y (2015) Coupled marginal discriminant mappings for low-resolution face recognition. Optik-International Journal for Light and Electron Optics 126:4352–4357
99. Zhang Y, Lu Y, Wu H, Wen C, and Ge C (2016) Face occlusion detection using cascaded convolutional neural network, in Chinese Conference on Biometric Recognition, pp. 720–727.
100. Zhang D-x, An P, Zhang H-x (2018) Application of robust face recognition in video surveillance systems. Optoelectron Lett 14:152–155
101. Zhang MM, Shang K, Wu H (2019) Learning deep discriminative face features by customized weighted constraint. Nuerocomputing 332:71–79
102. Zhao S (2018) Pixel-level occlusion detection based on sparse representation for face recognition. Optik 168:920–930

103. Zhao K, Jingyl X, and Cheng MM (2019) Regukarface: Deep face recognition via exclusive regularization", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1136–1144.
104. Zhou H, Lam K-M (2018) Age-invariant face recognition based on identity inference from appearance age. Pattern Recogn 76:191–202
105. Zhou Z, Wagner A, Mobahi H, Wright J, and Ma Y (2009) Face recognition with contiguous occlusion using markov random fields, in Computer Vision, 2009 IEEE 12th International Conference on, pp. 1050–1057.
106. Zhou L-F, Du Y-W, Li W-S, Mi J-X, Luan X (2018) Pose-robust face recognition with Huffman-LBP enhanced by divide-and-rule strategy. Pattern Recogn
107. Zhou Q, Zhang C, Yu W, Fan Y, Zhu H, Xiaofu W (2018) Face recognition via fast dense correspondence. Multimed Tools Appl 77:10501–10519
108. Zhuang L, Chan T-H, Yang AY, Sastry SS, Ma Y (2015) Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment. Int J Comput Vis 114:272–287

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