Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 03, September 2021.

COMPARATIVE STUDY OF MACHINE LEARNING AND DEEP LEARNING ALGORITHM FOR FACE RECOGNITION

Nikita Singhal[1], Vaishali Ganganwar[2], Menka Yadav[3], Asha Chauhan[4], Mahender Jakhar[5] and Kareena Sharma[6]

(Received: 28-Jun.-2021, Revised: 17-Aug.-2021, Accepted: 19-Aug.-2021)

ABSTRACT

In the present world, biometric systems are used to analyze and verify a person's distinctive bodily or behavioral features for authentication or recognition. Till now, there are numerous authentication systems that use iris, fingerprint and face feature for identification and verification, where the face recognition-based systems are most widely preferred, as they do not require user help every time, are more automated and are easy to function. This review paper provides a comparative study between various face recognition techniques and their hybrid combinations. The most commonly used datasets in this domain are also analyzed and reviewed. We have also highlighted the future scope and challenges in this domain, as well as various Deep Learning (DL)-based algorithms for facial recognition.

KEYWORDS

Face recognition, Local binary pattern, Convolutional neural networks, Principal component analysis, Histogram of oriented gradient.

1. INTRODUCTION

With the evolution of humans in every field of technology, there is a need to control who can access the place, machinery or information; so, we require an authentication system. There are many human authentication systems, such as signature, password, pin and biometric systems that have been developed. Face authentication systems have become popular as they doesn’t disturb the privacy of the individual and there is no requirement to get in physical contact with the system, which helps in controlling the spread of diseases like viruses. Face authentication is defined as giving access to the authorized person; i.e., face identification problem. It is a two-step process; firstly face detection, which is the detection of the human face in the frame of the image or video and highlighting it by making a square around the face discarding the surrounding and secondly Face Recognition (FR), which means the face detected in the above step has to be verified with those present in the database and if there exists a match, then the person is authorized by the system; if not, then the owner can take the necessary measures. There are many factors the affect the FR algorithm, including physical factors (e.g. illumination, occlusion) as well as facial features (e.g. twins, relatives, pose and aging factor). The methods addressing all these issues have been surveyed in [1] by Mortezaie et al. To achieve the best results for FR, we also require expertise in the subject of psychology, so that we can study the feature characteristics of the face. Lots of work has been done on the FR from the standard algorithms, like Principal Component Analysis (PCA), Local Binary Pattern (LBP) to the latest DL methods, like Convolutional Neural Networks (CNNs).

The organization of the paper is as follows. In Section 2, we provide the main steps involved in the process of FR. In Section 3, we summarize the various FR algorithms based on ML and DL. In Section 4, we provide open challenges and directions for future scope and in section 5, we conclude the work.

2. STEPS INVOLVED IN THE PROCESS OF FR

FR can be considered as a way of authentication and verification. In this sence, a new unknown face is matched with various other faces present in the database which all have known entities. After this

N. Singhal, V. Ganganwar, M. Yadav, A. Chauhan, M. Jakhar and K. Sharma are with the Department of Computer Engineering, Army Institute of Technology, Pune, India. Emails: [[1]ngupta, [2]vganganwar]@aitpune.edu.in, [[3]yadavmenka2009, [4]ashachauhan8085, [5]jakharmahender8 and [6]karinasharma1119]@gmail.com)]

comparison, a result is given out signifying whether the face has been recognized or not. The face identity is either confirmed or denied by the result drawn after comparison with the face data present in the database. FR process consists of two major components to carry out the whole process that is face detection and FR.

2.1 Face Detection

It involves detecting a face or all faces in a given image or video by using various detection techniques. Its robustness to pose, illumination and the elimination of background results in better detection of faces.

The Viola-Jones [2] is the most used face detector that is based on Haar-like features and shows better results for front faces in its real-time implementation. Some deep learning-based methods are also used for detecting faces, like the sliding-window idea [3], R-CNN [4] and the single-shot detector (SSD) [5] that are successfully used for face detection and provide good results.

2.2 Face Recognition

FR is one of the crucial parts where the detected face after conversion in grayscale is recognized and compared with certain images for authentication or identification. It takes the image detected as input and then checks it with the images present in the database for validation.

The detected face is compared against the database features and if there is a match, then the face is recognized properly and if not, authenticity is not provided to that user.

3. ALGORITHMS USED FOR FR

3.1 PCA-based

It is a statistical approach, where a set of possibly correlated variables' observations are converted into linearly uncorrelated variables' values known as principal component using orthogonal transformation. Here, data is transformed with help of its projection that is further treated as principal component for first coordinate and then such more variance is created called second component, … and so on resulting in a new coordinate system.

PCA has many advantages when applied to ML algorithms for factors like dimensionality reduction, feature transformation, data visualization as well as for Speeding up the machine learning algorithms and showing better results in terms of recognition rate compared to other techniques, especially to recognize faces with expression disturbance and background disturbance.

PCA and its combination with other algorithms have been widely used for FR application in the past few decades. Wang et al. [6] used LBP and PCA with ABAS algorithm combination for FR. The LBP and PCA were combined used as the feature extraction method and the ABAS algorithm was used for optimization of the neural network, while softmax function was used here to reduce the time for multi-face classification that was constructed to carry out the FR process. Here, the ORL [7] dataset was used to test the proposed model to showcase its capability to handle multi-face classification.

Wang et al. [8] used the F-2D-QPCA technique for FR. They used F-norm to maximize image variance and a greedy iterative algorithm for good convergence and robustness of the method. Experimental results of this model on several color face image databases showed its effectiveness and accuracy compared to other existing models, as their method uses an image as a quaternion matrix that uses color and spatial information of an image.

Kong et al. [9] proposed the CSGF (2D) 2PCANet algorithm for FR. The proposed model used CSGF to overcome the computation time and data redundancy problems of existing models. It consisted of one stage for non-linear output for which linear SVM was used and two stages for feature extraction that had good locality, for which two-dimensional PCA was used. The proposed model showed a higher recognition rate when it was tested on AR [10], ORL databases …etc. and had stable robustness to face image variations' resulting in improving the accuracy of the model.

Low et al. [11] presented a method to boost the performance of 2FGFC against other face descriptors by using the standard 40 multi-scale multi-orientation Gabor filters into the condensed Gabor filter ensemble of only 8 filters. The demodulated Gabor phase features are grasped by an average pooling operator followed by whitening PCA to obtain the final representation with better performance.

Zhang et al. [12] presented a reliable PCA-based FR outsourcing protocol. The proposed methods for privacy-preserving matrix addition, multiplication and vector multiplication ensured the safety of the used technique. The Freivalds algorithm was used here for result verification. The proposed method benefited in rapid development in FR while ensuring the security of the application.

Table 1 summarizes face recognition based on PCA and Figure 1 shows the performance of different PCA algorithms along with ML and DL methods on YALE B and AR databases.

Table 1. PCA-based methods for face recognition.

Study & Pub. Year	Method/Algorithm	Dataset	Accuracy
[6], 2020	Novel Multi-face Recognition, ABASNet	ORL, ExtYaleB [14] and FERET [15]	99.35%, 99.54% and 99.18%, respectively on three datasets
[8], 2020	A Quaternion PCA Method, F-2D-QPCA	GT [16], GT-noise, GT-outlier, FT and FT-outlier	F-2D-QPCA-72.29%, QRR-72.29%, QSR-71.86% (for 30 features)
[9], 2018	Deep Learning, CSGF(2D)^{2}PCANet	XM2VTS [17], ORL, Extend YaleB, LFW[18] and AR	99.58%, 97.50%, 100%, 98.58%, +97.50%, respectively on the five datasets
[13], 2017	Stacking-based CNNs, PCANet+	FERET, LFW and YTF [19]	94.23%

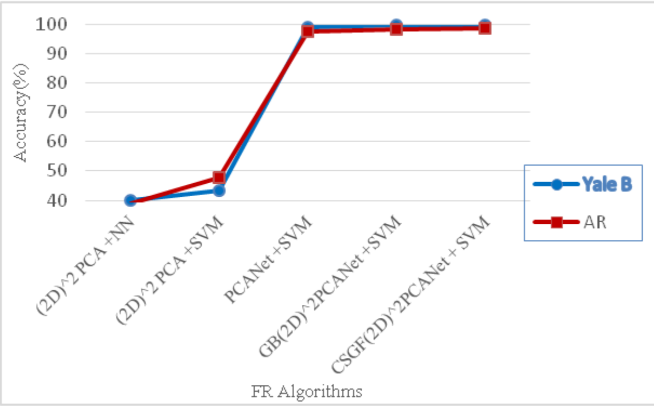

Figure 1. Accuracy of PCA along with ML and DL on YALE B and AR databases.

3.2 LBP-based

It is a texture operator that labels the image's pixels by thresholding each pixel's neighbourhood and using a binary number as a result to represent local features in images. It is considered to be useful for texture classification and improves the detection performance when used with histogram of oriented gradient (HOG).It is robust for monotonic grayscale transformations, gives a great result in a controlled environment and is one of the easiest FR algorithms.

Dalali et al. [20] used discrete wavelet transform as a preprocessing method for extracting significant features. Daubechies wavelets help in extracting approximation coefficients with single-level decomposition, so that the information for FR can be removed. The main focus is to reduce the information into a less significant coefficient, hence resulting in few storage uses. For this paper, the dataset taken into consideration was the MIT face dataset. Two types of performance results were obtained; i.e., for images with noise and for images without noise. For images without noise, an accuracy of 99.3% was achieved and for images with noise, the accuracy was 98.28%.

Tang et al. [21] used the LBP operator to extract features from the face texture and then used 10 CNNs with five different neural structures for more feature extraction for training purposes, as well as for network parameter improvisation and classification results by using the softmax function after the layers were fully connected. Finally, using majority voting, parallel ensemble learning was used to generate the final result of FR. The FR rate in the ORL was increased to 100% using this method, while Yale-B improved it to 97.51 percent.

To reduce the effect of face image variations on feature extraction performance, Muqeet et al. [22] proposed a method that uses directional wavelet transform (DIWT) and LBP to overcome the effect. The LBP histogram features were extracted from selected top-level DIWT sub-bands as a local descriptive feature set. The proposed method was tested on ORL, FEI [23] and GT databases. Results showed that the proposed method was more efficient than LGBP, LSPBPS and CTLBP methods.

Zhang et al. [24] proposed a face anti-spoofing strategy by using LBP, DWT (Discrete Wavelet Transform) and DCT (Discrete Cosine Transform) with an SVM classifier. In this paper, DWT-LBP features were generated that contained information regarding blocks' spatial details of video files and at last, the SVM classifier with RBF kernel was trained for anti-spoofing.

To improve the speed and accuracy of 3-dimensional FR, Shi et al. [25] presented an SVM and LBP combination.LBP was used for feature extraction of details of the 3-D face depth image. After that, information classification was done by SVM. The databases used for experimentation were the Texas 3DFRD 3-D face depth database and self-made depth database. The results showed that the proposed method had a low time consumption and a high recognition rate. Table 2 summarizes face recognition based on LBP.

Table 2. LBP-based methods for face recognition.

Study& P. Year	Method/Algorithm	Dataset	Accuracy
[20], 2016	LBP	MIT face database	Without noise: Max. 99.3%, Min. 99.0% With noise: Max. 98.28%, Min. 97.82%
[21], 2020	CNN and LBP	ORL and Yale-B	ORL: 100%, Yale-B: 97.5%
[22], 2017	LBP	ORL , GT and FEI	ORL: 97%, GT: 82.25%, FEI: 91.14%
[24], 2020	DWT-LBP-DCT	REPLAY-ATTACK [26] and CASIA-FASD [27]	REPLAY-ATTACK: 7.361%, CASIA-FASD: 93.84%
[25], 2020	LBP and SVM	Texas 3DFRD3-D face depth and self-made depth databases.	96.83%

3.3 HOG-based

It is a simple feature descriptor used in image processing to extract features from images and to detect objects. A feature descriptor simplifies an image by extracting only the information that is needed and discarding the rest.HOG features are useful for the first step in detecting objects. Gradient-based representation is obtained from pixel-based representation and is used with linear classification techniques and multi-scale pyramids for object detection.

Zemgulys et al. [28] proposed an image segmentation method using HOG and SVM algorithms for classification. Two approaches were discussed to detect the hand gestures of the referee; i.e., the wearable sensors and computer vision to recognize the signals from the referee in a basketball match, where an accuracy of 97.5% was achieved and the F1-score was 94.95%.

Rameswari et al. [29] implemented an access control system where face detection and recognition were the main parameters for the access control and the HOG was used for feature extraction and facenet algorithm for FR. Along with that, FR RFID technology was used to make the system more secure. In this system, the FaceNet algorithm achieved a higher accuracy of 97% compared other face detection algorithms, like LBPH, FisherFace, …etc.

Chitlangia et al. [30] proposed a method in which the personality trait of an individual is predicted based upon his/her handwriting. HOG was used for feature extraction and those features acted as an input to the SVM model, where classification of the writer's personality traits was done into Introvert, Optimistic, Energetic, Sloppy and Extrovert. The proposed method using the polynomial kernel had 80% accuracy.

Lakshmi et al. [31] used LBP with modified HOG features for facial expression recognition and a multi-class SVM algorithm was used for classification and recognition. Two datasets used were JAFFE [32] and CK+ [33] datasets. The accuracy with CK+ dataset was 97.66%.

Yan et al. [34] used HOG, Adaboost and SVM combinations for the application of real-time vehicle detection. The HOG was used for feature extraction and then the AdaBoost classifier was trained by the combination of HOG features and the dataset that is used in Treatment Group of Images for the classifier training. The accuracy with the HOG and AdaBoost combination was 97.24% and it reached 96.89% using the HOG features with the SVM classifier. Table 3 summarizes face recognition based on HoG.

Table 3. HoG-based methods for face recognition.

Study& P. Year	Method/Algorithm	Dataset	Accuracy
[28], 2018	HOG and SVM	Private	97.5%
[29], 2020	HOG and FaceNet	Private	97%
[30], 2019	HOG and SVM	Private	80%
[31], 2021	Modified HOG along with LBP and SVM	JAFFE and CK+	JAFFE: 90.83% and CK+ : 97.66%
[34], 2016	HOG with AdaBoost and SVM classifier	GTI vehicle [35]	AdaBoost: 97.24% and SVM: 96.89%

3.4 SVM-based

SVM provides a new dimensionality to pattern recognition problems. It can solve face recognition problems with both linear and nonlinear SVM training models. As it requires less computation power, it is commonly used in ML classification problems.

Zhang et al. [36] have extracted multi-scale features from the images of 20 subjects each having different poses with seven different expressions by using bi-orthogonal wavelet entropy to extract multi-scale features. They also employed a strict validation model using stratified cross-validation. They have achieved results superior to three state-of-the-art methods with the accuracy of 96.77% using fuzzy multiclass support vector machines to be classifiers.

The main aspect, according to Pham et al. [37], is to overcome the problem encountered in CNNs when we have imbalanced training data points for classes by increasing the number of training samples of the minority class. They created an image with similar facial expressions using the Action Units (AU) feature set. To improve the model's overall efficiency, AU features are combined with CNN features to train SVM for classification.

Omara et al. [38] developed multimodal biometric systems using hybrid Learning Distance Metric and Directed Acyclic Graph SVM models. The model was tested on an AR face dataset and achieved an accuracy of 99.85%, outperforming many state-of-the-art multi-modal methods. Kernel SVM is used as a classifier which provided better results than traditional classifiers.

Zhang et al. [39] detected athletes' fatigue states by developing an SVM-based model keeping the acceptance criteria of the Sequential Forward Floating Selection (SFFS) algorithm. They used an adaptive median filter method to remove noise and smooth the image and an adaptive threshold light equalization method to adjust the light. The dimensionality of the entire feature set is reduced and a fatigue motion feature subset is extracted. If the face images have more than 80% of their eyes closed, the method classifies them as fatigued. Table 4 summarizes face recognition based on SVM with different techniques.

Table 4. SVM-based methods for face recognition.

Study& P. Year	Method/Algorithm	Dataset	Accuracy
[36], 2016	Fuzzy SVM and Stratified Cross-validation	20 subjects X 7 different expressions (Private)	96.77+-0.10%
[37], 2019	SVM fused with CNN (DenseNet) and AU features	RAF, Fer2013, ExpW	RAF: 91.37%, Fer2013: 71.01%, ExpW: 72.84%
[38], 2021	Distance Metric and DAG SVM	AR face dataset	99.85%
[39], 2020	SVM and SFFS	8000 face images (Private)	Above 90%

3.5 CNN-based

In recent years, CNNs have been recommended to solve computer vision problems as they have shown tremendous growth. The convolution and pooling layers of the CNN can extract the maximum amount of facial features compared to standard algorithms when used for FR. As the amount of training data is increasing in this digital world, we need a deep learning model that takes significantly less amount of time to train the model.

Syafeeza et al. [40] challenged the main factors (illumination variances, poses, facial expressions, occlusions) which affect the performance of the face recognition algorithm by proposing a robust 4-layer CNN architecture. The system achieved an accuracy of 99.5% on AR database and 85.13% on FERET database (on its 35 subjects). The most significant feature of the system was that it takes less than 0.01 second to complete the FR process.

Zangeneh et al. [41] used a coupled mapping method architecture for high- and low-resolution face images that have two branches of deep convolutional neural networks for each type of resolution to be converted into a common space. The branch associated with the conversion to the common space from high-resolution consists of fourteen-layer network, whereas the branch corresponding to low-resolution face image transformation consists of an additional network of 5 layers that was connected to the 14-layer network. It was tested on FERET, LFW and MBGC [42] datasets, where the proposed architecture proved a 5% better accuracy that is 97.2% compared to the traditional methods implemented before and outperformed the other methods, showing good performance when applied to very low-resolution images of 6*6 pixels.

Im et al. [43] proposed an authentication system for preserving the privacy of Smartphone users against malicious clients by storing a feature vector of the face in the encrypted form. Euclidean distance-based matching score is computed whenever someone tries to access the private vector on the remote server. It takes 1.3 seconds to perform the secure face verification in real-time whereas it takes just 1 second for the CFP [44] and ORL datasets to face verification. To further improve the computational score, they used the Catalano-Fiore transformation that converts a linear homomorphic encryption scheme into a quadratic scheme.

Goel et al. [45] have used a high-level method of feature extraction based on the DCNN-Optimized Kernel Extreme Learning Machine algorithm. Particle Swarm Optimization (PSO) algorithm is used for parameter optimization alongside polynomial function Kernel ELM classification algorithm. The results achieved without normalization on the datasets AT&T, CMU-PIE [46], Yale [47] and UMIST [48] were 0.5, 8.89, 0 & 21 error rate. This method has the least training time compared to other DLNs.

Zhao et al. [49] handled various face presentation attacks by proposing a deep architecture to increase the accuracy of multi-view human FR. Here, the authors proposed a CNN for extracting face features and to further localize the key points on the face, it has used the face alignment algorithm. PCA was used for dimensionality reduction of the deep features and a joint Bayesian framework (JBF) was proposed to score the similarity of feature vectors. An accuracy of 98.52% was achieved on CAS-PEAL [50] dataset.

To address the challenge of automatic age estimate in real-time applications, Al-Shannaq et al. [51] proposed a model for estimating human age using a fine-tuned CNN model. Two types of datasets were used to evaluate the idea. The MAE for the FG NET (limited) dataset was 3.446, while the MAE for the UTKFace (unconstrained) dataset was 4.867. Using the Adience dataset, the model was fine-tuned for the age group classification task and the overall accuracy the model achieved was 61.4%. Table 5 summarizes face recognition based on CNN.

Table 5. CNN-based methods for face recognition.

Study& P. year	Method/Algorithm	Dataset used	Accuracy
[40], 2014	4-layer CNN architecture	AR, FERET	AR: 99.5% and FERET: 85.13%
[41], 2019	Coupled mapping method, DCNNs	FERET, LFW and MBGC	FERET: 99.2%, LFW: 76.3% and MBGC: 68.64%
[43], 2020	Euclidean distance-based, Catalano-Fiore transformation	CFP,ORL	EER: 1.17 ,0.37

[45], 2020	OKELM algorithm, PSO, polynomial function KELM	AT&T,CMU-PIE,YALE,UMIST	EER: 0,0,6.67,10.9
[49], 2020	CNN + PCA,JBF	CAS-PEAL	98.52%
[51], 2020	CNN	FG NET, UTKFace, Adience	MAE: 3.446, MAE: 4.867, 61.4%

3.6 AlexNet-based

The name AlexNet refers to a CNN that has a significant impact in the field of DL for computer vision. It comprises of data augmentation, (1111, 55, 33, convolutions), dropout, max pooling, ReLU activations and SGD with momentum.

Suleman Khan et al. [52] proposed an advanced smart-glasses' framework capable of recognizing faces. The use of portable smart glasses to implement facial recognition can assist law-enforcement officials in recognizing a suspect's face. They have an advantage over security cameras due to their portability and superior frontal view capturing. This technique has a detection rate of 98 % when using 3099 features. AlexNet is used for facial recognition and after training 2500 photos in each class, it has achieved 98.5 % accuracy. During recognition, problems such as emotions and light intensity can be overcome by using a large number of different photos.

Hailong Yu et al. [53] proposed a method in which feature extraction is improved by employing an MLP convolutional layer. The CASIA-Web dataset is used for training and testing. After 10575 trials, the model has achieved 82.3% identification rate. For face verification, the LFW face database was used and 6000 pairs of face comparison trials were calculated, yielding an average recognition rate of 84.5%.

Suleman Khan et al. [54] proposed a framework for facial recognition based on AlexNet and transfer learning. This network requires a vast database to train, but the accuracy is great. They used four different classes from the database for training, with 1000 photos in each class and achieved an accuracy of 97.95%. Table 6 summarizes face recognition based on AlexNet.

Table 6. AlexNet-based methods for face recognition.

Study& Publication year	Method/Algorithm	Dataset used	Accuracy
[52], 2019	CNN + AlexNet	2500 variant images in a class using 3099 features	98.5%
[53], 2019	MLP + MFM in CNN	CASIA-Web data LFW face database	82.3% 84.5%
[54], 2019	AlexNet	1000 different people	97.95%

3.7 ResNet-based

A residual neural network (ResNet) is a popular deep learning model that uses residual blocks to overcome the problem of training extremely deep networks. It skips connections and leaps over layers by using these blocks. This skill facilitates the training of huge networks without increasing the percentage of training errors.

Ze Lu et al. [55] proposed a model for low-resolution FR called 'deep coupled ResNet model'. The trunk network, a ResNet-like network, was used to extract the discriminative features shared by face photos of various resolutions. Then, using branch networks, coupled mappings were learned to project features of images. The proposed model was experimented on LFW (with different probe sizes) and SCface datasets (with three different sets of dataset in accordance to camera distance), where it achieved 93.6% - 98.7% and 73.0% - 98.0% accuracy in face verification.

Storey et al. [56] introduced a 3DPalsyNet framework for mouth motion recognition and facial palsy grading in their research. For collecting the dynamic actions of the video data, they used a modified 3D CNN architecture with a ResNet backbone. The structure was tested using two datasets; CK+ and their own Facial Palsy dataset, achieving an F1-score of 82 % for mouth motion and 88% for facial palsy grading, respectively, where these values were greater than those for 3D CNN demonstrating its capacity to perform efficient facial analysis from video sequences.

Peng et al. [57] provided two approaches for FR. The first was to convert Inception-residual ResNet's scaling factor from a hyperparameter to a trainable parameter, with the initial tiny value of 0.1 ensuring a stable training network at the start. The second was to use Leaky ReLU and PReLU in the Inception-ResNet module, which boosts network performance by maximizing input data utilization. Both methods were tested on the VGGFace2, MS1MV2, IJBB and LFW datasets, with better accuracy and training process stability.

Li et al. [58] proposed an enhanced facial emotion recognition model by using ResNet-50 as the network backbone and CNN for feature extraction. To increase the model's convergence ability, BN and the activation function ReLU were used. The model was tested using their own dataset of 20 different subjects (700 images) with varying expressions and ages and it was found to have good accuracy of 95.39 ± 1.41%. Table 7 summarizes face recognition based on ResNet.

Table 7. ResNet-based methods for face recognition.

Study& P. year	Method/Algorithm	Dataset used	Accuracy
[55], 2018	Deep coupled ResNet model	LFW face database & SCface datasets	93.6 - 98.7 % 73.0 - 98.0 %
[56], 2019	Modified 3D CNN + ResNet	CK+ dataset	F1 score: 82%
[57], 2019	Modified 3D CNN + ResNet	Private Facial Palsy dataset	F1 score: 88%
[58], 2021	CNN + ResNet-50	Private dataset of 20 different subjects (700 images)	95.39 ± 1.41%

3.8 Comparison of All Methods

The performances of some FR algorithms which are using commonly used LFW and ORL datasets are shown in Figure 2 and Figure 3. In Figure 2, we compared CSGF(2D)2PCANet [9] with a Linear SVM approach PCANet+ [13] with DL-based AlexNet [53] and ResNet [55] model on LFW as common database, where CSGF(2D)2PCANet was proven to be better with an accuracy rate of 98.58% compared to other models.

Figure 3 shows the comparison of the performances of these algorithms on ORL face dataset. PCANet with a Linear SVM approach [9] yielded 91% accuracy and (2D) PCA with NN in [9] yielded approximately 97% accuracy, CNN with LBP combination [21] gave 100% accuracy, HOG + LDP with linear SVM [59] achieved approximately 97% accuracy and the simple HOG with linear SVM model [59] gave approximately 90% accuracy over this dataset, whereas for CNN, we studied ESPCN with CNN combination [60], which gave approximately 93% accuracy over this dataset. Based on the results, we concluded that the LBP and HoG combination with ML had higher accuracy and performed better than other models on the ORL dataset.

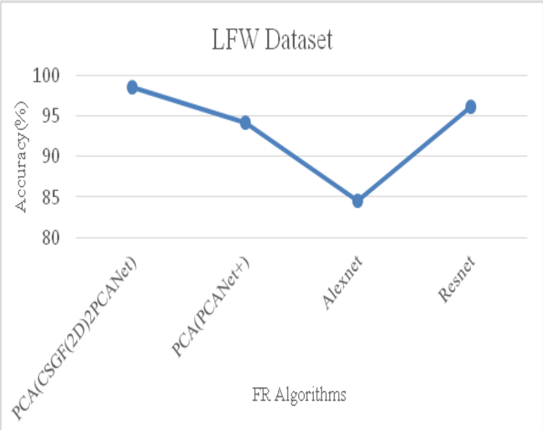

Figure 2. Performance of FR algorithms on LFW database.

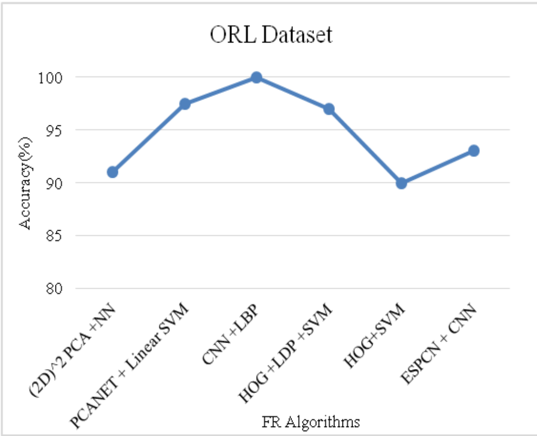

Figure 3. Performance of FR algorithms on ORL database.

3.9 Dataset

Table 8 summarizes the popular datasets used by researchers for FR.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No. 03, September 2021.

Table 8. Summary of datasets used for FR.

Year	Dataset	Total Images/Videos	Features
1994	ORL [7]	400 images	Various facial expressions, facial details (glasses or no glasses) and lighting conditions were used in the images. With the subjects in a frontal, upright position, a dark, homogenous background was used.
1997	YALE [47]	165 grayscale images	Each subject has 11 photos, one for each different face emotion or configuration: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, drowsy, shocked and wink.
1998	AR [10]	4000 images, 126 persons (56 women and 70 men)	Images consist of different expression (i.e., Neutral expression, Smile, Anger, Scream), various lighting conditions, with and without wearing sunglasses or a scarf.
1998	FERET [15]	14,126 images, 1199 individuals	It consists of 1564 sets of images with 365 duplicate sets of images.
1998	JAFFE [32]	213images	Ten Japanese female models were used to pose for 7 facial expressions (6 fundamental face expressions + 1 neutral).
1998	UMIST [48]	564 images, 20 subjects	Every image features a variety of positions, ranging from profile to frontal views. Subjects represent a diverse spectrum of races and genders resulting in a more comprehensive dataset.
1999	XM2VTS [17]	2360 mug shots, 295 individuals	Dataset is supplied with manually located eye points for all 2360 images for better recognition.
2001	YALE-B [14]	5760 images	Each subject is viewed in 576 different ways (9 poses x 64 illumination conditions).A photograph with ambient (background) illumination was also captured for each individual in a certain stance.
2001	Extend Yale B [14]	2414 images, 38 subjects	Images were taken in a variety of lighting circumstances and with a variety of facial expressions, resulting in an excellent result.
2002	CMU-PIE [46]	41,368 images	Each photograph was taken in 13 distinct stances, with 43 various lighting situations and four distinct expressions.
2003	CAS-PEAL [50]	99594 images	Chinese face database with large-scale images. To gather 27 photos in three shots, each individual is instructed to gaze straight ahead, up and down. The database also includes five facial expressions, six accessories and 15 lighting adjustments.
2007	LFW [18]	13,233 images, 1680 people	Its goal is to collect facial images and other relevant data for Wikipedia's Living People category.
2009	MBGC [42]	628 Videos	There are 4025 frames in which the left iris is visible and 4013 frames in which the right iris is visible. An Iris On the Move (IOM) technology took the near-infrared facial video.
2009	Multi-PIE [61]	750,000 images, 337 people	15 views and 19 lighting settings were used to photograph the subjects with various expressions and frontal images with high resolution.
2010	FEI [23]	2800 images	Faces of people between the ages of 19 and 40, each having a distinct appearance, haircut and adornments were used for images.
2011	YTF [19]	3,425 videos, 1,595 persons	Used YouTube as a source for videos. Each subject has an average of 2.15 videos available. The average length of a video is 181.3 frames with the shortest clip of 48 frames and the longest clip of 6,070 frames.
2012	CASIA-FASD [27]	600 (240 for training and 360 for testing), 50 subjects (12 videos per subject)	Anti-spoofing dataset. Videos were taken under different light conditions and resolutions.
2012	GTI-Vehicle [35]	3425 images of vehicle rears, 3900 images extracted from road sequence	This database has images extracted from a video sequence. The images cover different driving conditions, especially related to weather.

2012	Replay-Attack [26]	1300 videos	It is a face-spoofing database. All videos were created by displaying a snapshot or video recording of the same client for at least 9 seconds or having an actual client try to access a laptop through a built-in webcam.
2016	CFP [44]	500 individuals (each subject has10 frontal and 4 profile images)	It has 10defined splits, each containing 350 same and 350 not-same pairs.

4. Open Challenges for Future Research

The early dataset that was used consisted of images taken under specified and controlled environments. The accuracy of the algorithm is severely affected under adverse conditions of the image, such as low resolution, blur, pose variation and occlusion. Most of the image-based data available today is obtained using low-resolution devices and to get higher accuracy with this data is a challenge. The latest huge datasets created using images from the internet are not annotated properly, which results in poor accuracy of the DNN models. They are also prone to face-spoofing attacks as they can be easily deceived. So, there is a requirement for more robust DNNs. Video-based datasets yield better results as we can capture the dynamic aspects of the face which helps counter spoofing attacks on the networks.

5. Conclusions

In this paper, we explored existing FR techniques based on various descriptor methods combined with machine learning classifiers, such as SVM, deep learning and transfer learning. We also listed the popular datasets used for FR technique. The limitations of existing FR technique are that, they used the datasets of images taken under specified and controlled environments and the performance of these systems degrades under adverse conditions known as semantic adversarial attacks or when downloaded from the internet. Our study will provide insight into existing FR techniques for researchers who wish to conduct their research in this field. The challenge for the future study is to develop a robust FR algorithm that can handle low-resolution images captured in an uncontrolled environment.

References

[1] Z. Mortezaie and H. Hassanpour, "A Survey on Age-invariant Face Recognition Methods," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 05, no. 02, pp. 87-96, August 2019.

[2] W. LU and M. YANG, "Face Detection Based on Viola-Jones Algorithm Applying Composite Features," Proc. of the IEEE Int. Conf. on Robots & Intell. Sys. (ICRIS), pp. 82-85, Haikou, China, 2019.

[3] H. Li, Z. Lin, X. Shen, J. Brandt and G. Hua, "A Convolutional Neural Network Cascade for Face Detection," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 5325-5334, DOI: 10.1109/CVPR.2015.7299170, Boston, USA, 2015.

[4] H. Chen, Y. Chen, X. Tian and R. Jiang, "A Cascade Face Spoofing Detector Based on Face Anti-spoofing R-CNN and Improved Retinex LBP," IEEE Access, vol. 7, pp. 170116-170133, 2019.

[5] X. Hu and B. Huang, "Face Detection Based on SSD and CamShift," Proc. of the IEEE 9th Joint Int. Inform. Techn. and Artificial Intelligence Conf. (ITAIC), pp. 2324-2328, Chongqing, China, 2020.

[6] F. Wang, F. Xie, S. Shen, L. Huang, R. Sun and J. Le Yang, "A Novel Multi Face Recognition Method with Short Training Time and Lightweight Based on ABASNet and H-Softmax," IEEE Access, vol. 8, pp. 175370-175384, DOI: 10.1109/ACCESS.2020.3026421, 2020.

[7] F. S. Samaria and A. C. Harter, "Parameterization of a Stochastic Model for Human Face Identification," Proc. of IEEE Workshop on Applications of Computer Vision, 1994, pp. 138-142, Sarasota, USA, 1994.

[8] M. Wang, L. Song, K. Sun and Z. Jia, "F-2D-QPCA: A Quaternion Principal Component Analysis Method for Color Face Recognition," IEEE Access, vol. 8, pp. 217437-217446, 2020.

[9] J. Kong, M. Chen, M. Jiang, J. Sun and J. Hou, "Face Recognition Based on CSGF(2D)2PCANet," IEEE Access, vol. 6, pp. 45153-45165, DOI: 10.1109/ACCESS.2018.2865425, 2018.

[10] A. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report 24, [Online], Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html, 1998.

[11] C. Low, A. B. Teoh and C. Ng, "Multi-fold Gabor, PCA and ICA Filter Convolution Descriptor for Face Recognition," IEEE Trans. on Circuits and Systems for Video Techn., vol. 29, no. 1, pp. 115-129, 2019.

[12] Y. Zhang, X. Xiao, L. Yang, Y. Xiang and S. Zhong, "Secure and Efficient Outsourcing of PCA-based Face Recognition," IEEE Trans. on Information Forensics and Security, vol. 15, pp. 1683-1695, 2020.

[13] C. Y. Low, A. B. J. Teoh and K. A. Toh, "Stacking PCA Net+: An Overly Simplified Conv Nets Baseline for Face Recognition," IEEE Signal Processing Letters, vol. 24, no. 11, 2017.

[14] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643–660, 2021.

[15] P. J. Phillips, H. Wechsler, J. Huang and P. J. Rauss, "The FERET Database and Evaluation Procedure for Face-recognition Algorithms," Image and Vision Computing, vol. 16, no. 5, pp. 295-306, 1998.

[16] The Georgia Tech Face Database, [Online], Available: http://www.anefian.com/research/face_reco.htm.

[17] K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," Proc. of the 2nd International Conference on Audio and Video-based Biometric Person Authentication, vol. 964, pp. 965-966, 2000.

[18] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Proc. of Workshop on Faces in 'Real-Life' Images: Detection, Alignment and Recognition, [Onlive], Available: https://hal.inria.fr/inria-00321923, Marseille, France, 2008.

[19] L. Wolf, T. Hassner and I. Maoz, "Face Recognition in Unconstrained Videos with Matched Background Similarity," Proc. of CVPR 2011, pp. 529-534, DOI: 10.1109/CVPR.2011.5995566, Colorado Springs, USA, 2011.

[20] S. Dalali and L. Suresh, "Daubechives Wavelet Based Face Recognition Using Modified LBP," Procedia Computer Science, vol. 93, pp. 344-350, 2016, DOI:10.1016/j.procs.2016.07.219, 2016.

[21] J. Tang, Q. Su, B. Su, S. Fong, W. Cao and X. Gong, "Parallel Ensemble Learning of Convolutional Neural Networks and Local Binary Patterns for Face Recognition," Computer Methods and Programs in Biomedicine, vol. 197, DOI: 10.1016/j.cmpb.2020.105622, 2020.

[22] M. A. Muqeet and R. S. Holambe, "Local Binary Patterns Based on Directional Wavelet Transform for Expression and Pose-invariant Face Recognition," Applied Computing and Informatics, vol. 15, no.2, pp. 163-171, DOI: 10.1016/j.aci.2017.11.002, 2019.

[23] C. E. Thomaz and G. A. Giraldi, "A New Ranking Method for Principal Component Analysis and Its Application to Face Image Analysis," Image and Vision Computing, vol. 28, no. 6, pp. 902-913, DOI: 10.1016/j.imavis.2009.11.005, 2010.

[24] W. Zhang and S. Xiang, "Face Anti-spoofing Detection Based on DWT-LBP-DCT Features," Signal Processing: Image Communication, vol. 89, Paper ID: 115990, DOI: 10.1016/j.image.2020.115990, 2020.

[25] L. Shi, X. Wang and Y. Shen, "Research on 3D Face Recognition Method Based on LBP and SVM," Optik, vol. 220, Paper ID: 165157, DOI: 10.1016/j.ijleo.2020.165157, 2020.

[26] I. Chingovska, A. Anjos and S. Marcel, "On the Effectiveness of Local Binary Patterns in Face Anti-Spoofing," Proceedings of the IEEE International Conference of Biometrics Special Interest Group (BIOSIG), pp. 1-7, Darmstadt, Germany, 2012.

[27] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi and S. Z. Li, "A Face Anti Spoofing Database with Diverse Attacks," Proc. of the 5th IAPR IEEE International Conference on Biometrics (ICB), pp. 26-31, DOI: 10.1109/ICB.2012.6199754, New Delhi, India, 2012.

[28] J. Žemgulys, V. Raudonis, R. Maskeliūnas and R. Damaševičius, "Recognition of Basketball Referee Signals from Videos Using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM)," Procedia Computer Science, vol. 130, pp. 953-960, 2018.

[29] R. Rameswari, S. N. Kumar, M. A. Aananth and C. Deepak, "Automated Access Control System Using Face Recognition," Materials Today: Proceedings, vol. 45, DOI: 10.1016/j.matpr.2020.04.664, 2020.

[30] A. Chitlangia and G. Malathi, "Handwriting Analysis Based on Histogram of Oriented Gradient for Predicting Personality Traits Using SVM," Procedia Computer Science, vol. 165, pp. 384-390, DOI: 10.1016/j.procs.2020.0, 2019.

[31] D. Lakshmi and R. Ponnusamy, "Facial Emotion Recognition Using Modified HOG and LBP Features with Deep Stacked Autoencoders," Microprocessors and Microsystems, vol. 82, DOI: 10.1016/j.micpro.2021.103834, 2021.

[32] M. Lyons, M.Kamachi and J. Gyoba, "TheJapanese Female Facial Expression (JAFFE) Database," The Japanese Female Facial Expression (JAFFE) Dataset, Zenodo, DOI: 10.5281/zenodo.3451524, 1998.

[33] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-specified Expression," Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 94-101, DOI: 10.1109/CVPRW.2010.5543262, San Francisco, USA, 2010.

[34] G. Yan, M. Yu, Y. Yu and L. Fan, "Real-time Vehicle Detection Using Histograms of Oriented Gradients and AdaBoost Classification," Optik, vol. 127, no. 19, pp. 7941-7951, 2016.

[35] J. Arróspide, L. Salgado and M. Nieto, "Video Analysis Based Vehicle Detection and Tracking Using an MCMC Sampling Framework," EURASIP Journal on Advances in Signal Processing, vol. 2012, Article ID: 2012 (2), DOI: 10.1186/1687-6180-2012-2, 2012.

[36] Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Philips, Q.-M. Liu and S.-H. Wang, "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine and Stratified Cross Validation," IEEE Access, vol. 4, pp. 8375-8385, DOI: 10.1109/ACCESS.2016.2628407, 2016.

[37] T. T. D. Pham and C. S. Won, "Facial Action Units for Training Convolutional Neural Networks," IEEE Access, vol. 7, pp. 77816-77824, DOI: 10.1109/ACCESS.2019.2921241, 2019.

[38] I. Omara, A. Hagag, S. Chaib, G. Ma, F. E. Abd El-Samie and E. Song, "A Hybrid Model Combining Learning Distance Metric and DAG Support Vector Machine for Multimodal Biometric Recognition," IEEE Access, vol. 9, pp. 4784-4796, DOI: 10.1109/ACCESS.2020.3035110, 2021.

[39] F. Zhang and F. Wang, "Exercise Fatigue Detection Algorithm Based on Video Image Information Extraction," IEEE Access, vol. 8, pp. 199696-199709, DOI: 10.1109/ACCESS.2020.3023648, 2020.

[40] A. R. Syafeeza, M. Khalil-Hani, S. S. Liew and R. Bakhteri, "Convolutional Neural Network for Face Recognition with Pose and Illumination Variation," International Journal of Engineering and Technology (IJET), vol. 6, no. 1, pp. 44-57, 2014.

[41] E. Zangeneh, M. Rahmati and Y. Mohsenzadeh, "Low Resolution Face Recognition Using a Two-branch Deep Convolutional Neural Network Architecture," Expert Systems with Applications, vol. 139, Article ID: 112854, DOI: 10.1016/j.eswa.2019.112854, 2019.

[42] P. J. Phillips et al., "Overview of the Multiple Biometrics Grand Challenge," Proc. of Advances in Biometrics (ICB 2009), Part of Lecture Notes in Computer Science, vol. 5558. Springer, Berlin, Heidelberg, [Online], Available: https://doi.org/10.1007/978-3-642-01793-3_72, 2021.

[43] J. Im, S. Jeon and M. Lee, "Practical Privacy-Preserving Face Authentication for Smartphones Secure Against Malicious Clients," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2386-2401, DOI: 10.1109/TIFS.2020.2969513, 2020.

[44] S. Sengupta, J. Chen, C. Castillo, V. M. Patel, R. Chellappa and D. W. Jacobs, "Frontal to Profile Face Verification in the Wild," Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-9, DOI: 10.1109/WACV.2016.7477558, Lake Placid, USA, 2016.

[45] T. Goel and R. Murugan, "Classifier for Face Recognition Based on Deep Convolutional Optimized Kernel Extreme Learning Machine," Computers & Electrical Engineering, vol. 85, Paper ID: 106640, DOI: 10.1016/j.compeleceng.2020.106640, 2020.

[46] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination and Expression Database," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp. 1615-1618, Dec. 2003.

[47] A. Georghiades, P. Belhumeur and D. Kriegman, "Yale Face Database," Yale University, [Online] Available: http://cvc.yale.edu/projects/yalefaces/yalefa, 1997.

[48] F. Zhao, J. Li, L. Zhang, Z. Li and S. Na, "Multi-view Face Recognition Using Deep Neural Networks," Future Generation Computer Systems, vol. 111, pp. 375-380, DOI: 10.1016/j.future.2020.05.002, 2020.

[49] W. Gao et al., "The CAS-PEAL Large-scale Chinese Face Database and Baseline Evaluations," IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, vol. 38, no. 1, pp. 149-161, DOI: 10.1109/TSMCA.2007.909557, 2008.

[50] A. Al-Shannaq and L. Elrefaei, "Age Estimation Using Specific Domain Transfer Learning," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 06, no. 02, pp. 122-139, June 2020.

[51] S. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah and S. U. Ali, "Facial Recognition Using Convolutional Neural Networks and Implementation on Smart Glasses," Proc. of the IEEE International Conference on Information Science and Communication Technology (ICISCT), pp. 1-6, Karachi, Pakistan, 2019.

[52] Z. Ren and X. Xue, "Research on Multi Pose Facial Feature Recognition Based on Deep Learning," Proc. of the 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), pp. 1427-1433, DOI: 10.1109/ICMCCE51767.2020.00313, Harbin, China, 2020.

[53] S. Khan, E. Ahmed, M. H. Javed, S. A. A Shah and S. U. Ali, "Transfer Learning of a Neural Network Using Deep Learning to Perform Face Recognition," Proc. of the International Conference on Electrical, Communication and Computer Engineering (ICECCE), pp. 1-5, Swat, Pakistan, 2019.

[54] Z. Lu, X. Jiang and A. Kot, "Deep Coupled ResNet for Low-resolution Face Recognition," IEEE Signal Processing Letters, vol. 25, no. 4, pp. 526-530, April 2018.

[55] G. Storey, R. Jiang, S. Keogh, A. Bouridane and C. Li, "DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework Using Fully 3D Convolutional Neural Networks," IEEE Access, vol. 7, pp. 121655-121664, DOI: 10.1109/ACCESS.2019.2937285, 2019.

[56] S. Peng, H. Huang, W. Chen, L. Zhang, W. Fang, "More Trainable Inception-ResNet for Face Recognition," Neurocomputing, vol. 411, pp. 9-19, DOI: 10.1016/j.neucom.2020.05.022, 2020.

[57] B. Li and D. Lima, "Facial Expression Recognition *via* ResNet-50," International Journal of Cognitive Computing in Engineering, vol. 2, pp. 57-64, DOI: 10.1016/j.ijcce.2021.02.002, 2021.

[58] H. Wang, D. Zhang and Z. Miao, "Fusion of LDB and HOG for Face Recognition," Proc. of the 37th IEEE Chinese Control Conf., pp. 9192-9196, DOI: 10.23919/ChiCC.2018.8483900, Wuhan, China, 2018.

[59] M. A. Talab, S. Awang and S. A. M. Najim, "Super-low Resolution Face Recognition Using Integrated Efficient Sub-pixel Convolutional Neural Network (ESPCN) and Convolutional Neural Network (CNN)," Proc. of the IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), pp. 331-335, DOI: 10.1109/I2CACIS.2019.8825083, Selangor, Malaysia, 2019.

[60] R. Gross, I. Matthews, J. Cohn, T. Kanade and S. Baker, "Multi-PIE," Image and Vision Computing, vol. 28, no.5, pp. 807-813, doi:10.1016/j.imavis.2009.08.002, 2010.

ملخص البحث:

في عالمنا الحاضر، تُستخدم أنظمة القياس الحيويّة لتحليل السِّمات المميّزة الجسديّة او السّلوكيّة لشخصٍ ما من أجل التّصديق أو التّمييز. وحتى الآن، هناك العديد من هذه الانظمة التي تستخدم العيْن أو بصمة الإصبع أو سِمات الوجْه للتّمييز بين الأشخاص والتّحقّق من هويّاتهم؛ إذ تُعتبر الأنظمة القائمة على تمييز الوجوه هي المفضّلة على نطاقٍ واسع، لأنها لا تتطلب مساعدة المُستخدِم في كل وقت، بالإضافة الى أنّها آليّة الى حدٍّ أكبر وسهلة التّشغيل.

هذا البحث عبارة عن ورقة مراجعة تقدّم دراسةً مقارنةً بين تقنيات متنوعة لتمييز الوجوه وتركيباتها الهجينة. كذلك تمّ تحليل أكثر مجموعات البيانات استخداماً في هذا الميدان ومراجعتها، الى جانب تسليط الضّوء على الآفاق المستقبلية والتحدّيات فيما يتعلق بموضوع البحث، والخوارزميات القائمة على التعلُّم العميق المستخدمة في مجال تمييز الوجوه.