Received 13 June 2024, accepted 2 July 2024, date of publication 8 July 2024, date of current version 14 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424933

SURVEY

A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges

H. L. GURURAJ[1], (Senior Member, IEEE), B. C. SOUNDARYA[1], S. PRIYA[1], J. SHREYAS[1], AND FRANCESCO FLAMMINI[2], (Senior Member, IEEE)

[1]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India

[2]IDSIA USI-SUPSI, University of Applied Sciences and Arts of Southern Switzerland, 6928 Manno, Switzerland

Corresponding authors: H. L. Gururaj (gururaj.hl@manipal.edu) and Francesco Flammini (francesco.flammini@mdu)

ABSTRACT Face Recognition (FR) is the technology used to identify and verify individuals based on their facial features. In recent decades, FR plays a crucial role in various sectors, including security, healthcare, banking, and criminal identification. Numerous techniques for effective FR are currently under development, ranging from appearance to hybrid approaches. Most of the existing methods offer diverse solutions to describe a face image either by focusing on specific facial features or by considering the entire face. This study explores a various range of such techniques and challenges related to FR. The existing solutions were analysed with respect to various perspectives of inputs, viz., illumination, pose variation, facial expressions, occlusions, and aging, which led to the prominent implementation of FR systems. The primary contribution of this survey lies in the comprehensive review of state-of-the-art FR techniques and deriving the taxonomy of categorizing these methods into various classes which range from appearance to hybrid approaches. Moreover, the proposed detailed study highlights the significant features used by the most recent research developed in FR, also, provide a detailed classification of image and video-based FR methods, highlighting major advancements and core processing steps for handling huge volume of datasets. Moreover, the proposed study outlines the current trends in available datasets and emphasizing their enhancements. This survey also aims to provide a valuable resource for researchers and practitioners by offering insights into the latest developments and identifying open problems that require further investigation.

INDEX TERMS Biometrics, face recognition, image processing, soft biometrics.

I. INTRODUCTION

In the modern era of digital world, proving the identity of every individual is highly essential. The traditional methods of identification which were of the form of ID card, driver's licenses, etc., are susceptible to theft, loss, or forgery. Moreover, individuals may misplace or have these physical documents stolen leads to identity theft or unauthorized access. In contrast, biometric authentication is inherent to an individual and cannot be easily replicated and can also provide secure means of identification. Among earlier biometric authentication which were in the form of fingerprint and iris scan, FR has become a crucial application in the field of biometric authentication systems [1]. Since, FR uses distinct facial features for detection, it provides a high level of accuracy and security in identifying individuals. With these prominent advantages, FR has become a crucial application in the field of biometric authentication systems over the past decades [2].

In recent years, many advancements came into picture to portray the successful integration of FR technique in numerous fields, viz., security systems, public safety systems, payment systems, information security, law enforcement and surveillance, smart cards, and access control [3], [4]. The applications of FR technology in different fields are given in Table 1. Though FR has several advantages, FR is a complex field that combines different areas of study viz., image processing, computer technology, machine learning, biology, and neural networks [5].

Generally, FR system undergoes two main phases such as i) Face verification and ii) Face identification. The first phase confirms a face matches a known identity, whereas

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyan Zhang.

TABLE 1. Exploring the diverse applications of FR technology.

Application Area	Description	Benefits	Challenges	Technological Innovations
Security and Surveillance	Access control systems, Video surveillance, Criminal identification, Border control	Enhanced security, Crime prevention, Efficient border management	Privacy concerns, Potential misuse	Facial recognition accuracy
Retail and Marketing	Customer analytics, Personalized marketing, Retail store surveillance for shoplifting prevention	targeted marketing, Shopper behavior analysis, Reduced retail theft	Shopper privacy concerns, Ethical implications	Automated cashier-less stores using facial recognition for payment processing
Healthcare	Patient identification, Medical record management, Monitoring	Accurate patient identification, Improved medication adherence	Patient data Security	AI-driven medical diagnosis and treatment recommendations based on facial analysis
Education	Student attendance tracking, -Exam security	Efficient attendance monitoring, Prevention of exam cheating	Student privacy concerns, System reliability	Facial recognition-based smart classroom attendance systems with real-time notifications to parents and teachers
Social Media	Photo tagging and organization, Friend suggestion algorithms	Safer online communities	User privacy concerns	Enhanced privacy controls with facial recognition-based account security settings
Airports and Travel	Boarding pass verification, Immigration and customs processing, Lost baggage tracking	Efficient immigration procedures, Reduced baggage mishandling	Traveler privacy concerns	Facial recognition-enabled baggage tracking with real-time updates to travelers

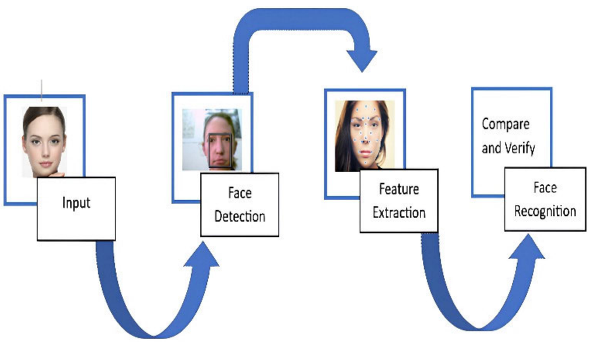

FIGURE 1. The workflow of a standard automated FR system biometric.

second phase used to figure out the matching face in the stored database. Face identification is a process where a system recognizes or identifies a person's identity by comparing their facial features with a database of known faces. This involves analysing various facial characteristics, viz., the shape of the eyes, nose, mouth, and the overall face structure to match with the stored data. Face verification is a process in which a system determines whether two facial images belong to the same person. It involves comparing the features extracted from the facial images, viz., facial shape and texture to confirm if they are similar enough to be considered the same person.

Figure 1 depicts the fundamental steps involved in developing a FR system such as i) face detection, ii) feature extraction, and iii) face recognition [6], [7]. In the face detection step, the system identifies and locates human faces in the images once they have been captured. The feature extraction step involves creating unique feature vectors for the detected faces. These vectors capture distinctive elements of captured face. Finally, in the face recognition step, the system compares the extracted features with a database of known faces and identifies the exact match in the stored templates. This is done by extracting distinctive facial features such as, position of the mouth, nose, eyes and other facial parameters. The other aspects of soft biometric systems rely on the usage of behavioral traits viz., signatures, walking patterns, speech, and facial dynamics in addition to static physiological traits like fingerprints, iris, and palm prints [3], [4].

The FR techniques can be classified into two-dimensional (2D) and three-dimensional (3D) based FR according to the kind of data used for recognition [8]. In literature, only limited researches have carried out on both 2D and 3D facial images. The rest of the research examined either 2D or 3D facial images, but not both [9].Though, 3D based FR offers several advantages [8], Most of the study prefer 2D based FR techniques due to its accessibility, affordability, and easier to implement compared to 3D based FR techniques. Moreover, the capturing devices used for 2D recognition are widely available and more cost effective than specialized 3D imaging devices.

From the literature reviewed it is evidenced that, the problem of FR is extensively studied in different research communities. The detailed study done in [10], provided a summary of recent advancements in 3D based FR which focus on three crucial aspects viz., pose, occlusion and expression recognition. To provide the clear picture of recent advancements in FR field, the Multi-Task Learning (MTL) for was introduced by [11]. In [12], [13], and [14], a concise review on human face detection techniques and future directions were presented. The review of numerous and extensive studies of FR techniques were widely discussed in [2] and [15] with different perspective. Reference [16] conducted comprehensive reviews of contemporary 3D based FR methods, delving into both traditional techniques and deep learning-based approaches. In [15], a thorough summary of the significant developments in deep FR, as well the advancements in learning facial representations for the

TABLE 2. Comparison of survey paper.

Reference	[3]	[15]	[10]	[12]	[11]	[21]	[108]	Our work
Detailed Survey	✓	X	✓	✓	✓	X	X	✓
Dataset Benchmark	X	✓	✓	X	✓	✓	✓	✓
Image-based	X	X	✓	X	✓	X	✓	✓
Video-based	X	X	X	X	X	X	X	✓
Review of face detection algorithm	✓	X	X	✓	✓	X	X	✓
Comparison	X	X	✓	✓	X	X	X	✓

purposes of verification and identification was proposed. The authors in [17] provided a comprehensive overview of facial expression analysis which encompassing RGB, 3D, thermal, and multimodal techniques. The Table 2 gives comparative analysis of various published papers based on several key points.

Hence, the main contribution of this study is to:

1. Provide an in-depth examination of various FR methods by analysing its each phase with respect to feature extraction, pre-processing, face detection, and classification.
2. Transcend the boundaries of still-image recognition to extend its focus to encompass video-based FR.
3. Provide the detailed analysis of various datasets used in 2D and 3D based FR techniques.
4. Address the applications and challenges in recent FR research.

The rest of the paper is organized as follows: Section II provides a background of steps involved in designing face recognition. In section III, a thorough analysis of various research frameworks initiated for both image-based and video-based FR techniques, including the contributions of deep learning (DL) methods and existing algorithms, is compared by considering the diverse variations such as pose, occlusion, and expression recognition. Moreover, the summary of publicly available benchmark datasets, tools, and evaluation metrics used for FR techniques are discussed in Section IV. The challenges and future research directions anticipated for FR techniques are given in Section V. Whereas the conclusion of the study is given in Section VI.

II. BACKGROUND

In this section, the core steps involved in designing the FR system are discussed in detail. Figure 1 illustrates the three fundamental steps in recognising face images, viz., (1) face detection, (2) feature extraction, and (3) face recognition. The details of each step are given below.

A. FACE DETECTION

The FR process starts by pinpointing the human face in an image. This initial step aims to determine the presence of human faces in the given picture. Nevertheless, challenges such as fluctuating lighting conditions and expressions can impede accurate face detection. To bolster the reliability of the FR system, specific pre-processing steps need to be implemented. Various techniques such as, Viola-Jones detector (VJ) [7], Histogram of Oriented Gradient (HOG) method [18], and Principal Component Analysis (PCA) [18], [19] were employed for detecting and locating human faces. The Viola-Jones face detector [7] is a widely used technique, especially effective for front-facing images which are operated in real-time and based on Haar-like features. These features are applied to images to capture essential elements such as, edges, corners, or lines which are fundamental to recognizing faces. Moreover, some other methods [20] incorporates data based on color to improve accuracy. These techniques collectively aid in identifying human faces and contribute to the robustness of the FR system.

B. FEATURE EXTRACTION

In this step, the goal is to extract features from the detected face images. The features can be extracted using either the global method or the local method. In the global method, the entire face will be focused. Whereas, in the local method the inner facial features or specific regions of interest will be focused [21]. Local methods are utilized to capture subtle information within specific facial regions. They are preferred because they are less influenced by factors like face geometry, aging, variations in pose, and face rotation [22], [23]. Further, the feature extraction methods can be classified into three main categories:

Generic Methods: These methods rely on identifying edges, lines, and curves in facial images. By considering measurements like size and distance, these techniques accurately identify and distinguish faces from one another [24]. To extract facial features, a variety of techniques are widely employed, including HOG [25], Eigenface [18], Independent Component analysis [ICA], Scale-Invariant Feature

Transform (SIFT) [26], Gabor filter [27], Local Phase Quantization (LPQ), Linear Discriminant Analysis (LDA) [24], ICA [28]. These methods play a crucial role in characterising and recognising facial attributes.

Feature-Template-Based Methods: These methods are designed to detect specific facial features, like eyes, using predefined templates. Generally, a face is represented using a set of features known as a signature. This signature describes the key facial features like the mouth, nose, and eyes, along with their geometric distribution within the face image [29]. The identification of each face is based on its distinctive structure, size, and shape [30].

Structural Matching Methods: These techniques consider geometrical constraints on facial features, ensuring they match specific structural patterns. The techniques such as Elastic bunch graph matching, Dynamic link architecture [31], and Local Binary Pattern (LBP) have shown prominent results in the literatures [31], [32], and [33].

C. FACE RECOGNITION(FR)

This is the final stage which recognises the identities of the face and enables automated FR. To accomplish this recognition, a face database is required, where the relevant facial data is stored for comparison and identification. For this purpose, multiple images of each individual are captured, and their distinctive features are extracted and stored in the database. When an input face image is submitted for recognition, it compares the extracted features with each face class stored in the database. In this step, the features extracted earlier and the facial elements are considered. These features are subsequently compared with the known faces stored in a specific database for recognition. Several techniques such as, Correlation Filters (CFs), Gaussian filter, median filter, Wiener filter, and histogram based on Peak Signal-to-Noise Ratio (PSNR) [33], Convolutional Neural Network (CNN) [34], K Nearest Neighbour (KNN), Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM) were used for detection. It is evidenced from the literature that the Gaussian filtering technique is highly efficient and ensures high-quality images without distortion [35].

III. TAXONOMY OF FACE RECOGNITION

In the realm of FR systems, existing literature categorizes these systems into two primary groups: i) Image based and ii) Video based methods. Image-based systems focus on recognizing individuals based solely on their physical appearance. In contrast, video-based systems not only consider physical features but also incorporate changes in appearance over time and dynamic facial movements. The general taxonomy of FR literature is illustrated in Figure 2, highlighting the distinction between these two fundamentals approaches

A. IMAGE-BASED FACE RECOGNITION

According to [6], Image based FR methods can be divided into three categories: i) Appearance methods, ii) Landmark methods and iii) Hybrid methods.

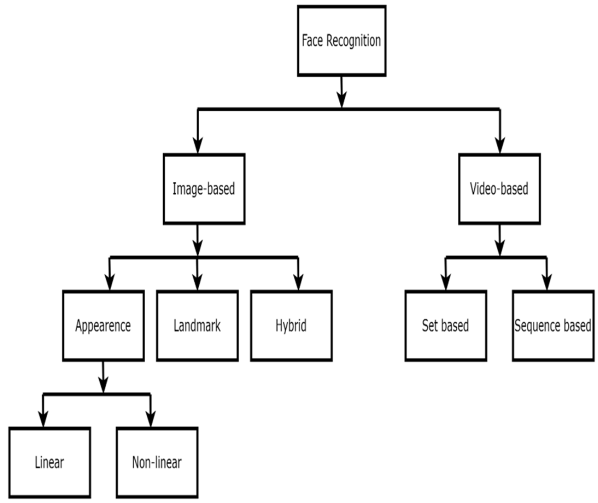

FIGURE 2. Taxonomy of face recognition.

1) APPEARANCE METHODS

The appearance approach, often called global feature-based methods works on the entire face to extract features. This approach aims to recognise faces by considering the entire facial representation, rather than focusing on individual components such as the mouth, eyes, or nose. These approaches primarily function by representing the face image as a matrix of pixels. This matrix is frequently converted into feature vectors, making it easier to process and analyse. Following this, the feature vectors are projected into a low-dimensional space. Nevertheless, both appearance and subspace techniques are sensitive to variations such as facial expressions, lighting, and poses. Despite this sensitivity, their advantages have led to widespread of use in FR applications. Furthermore, these approaches can be categorised into linear and non-linear techniques, depending on the method used to represent the subspace.

In the context of face recognition, a linear method is an approach that employs linear transformations or classifiers to analyse and recognise faces. These techniques usually include representing information in a dimensional space and using linear algebra methods for tasks, like reducing dimensions and classification. In literature, techniques such as PCA [36], Eigenface [37], LDA [37], and ICA [38] were used as prominent techniques in linear methods.

PCA: PCA is commonly employed to preprocess the data before further analysis. Reducing the dimensions in high-dimensional face data helps eliminate redundant information and noise. It preserves essential data characteristics, significantly lowers dimensionality, speeds up data processing, and ultimately saves time and costs. Hence, PCA is commonly employed for dimensionality reduction [39] and visualizing multi-dimensional data. It is especially efficient when dealing with large datasets [9]. PCA was used to represent global facial features in [40]. Likewise, [36] introduced the Modular Eigenspace description technique, incorporating prominent

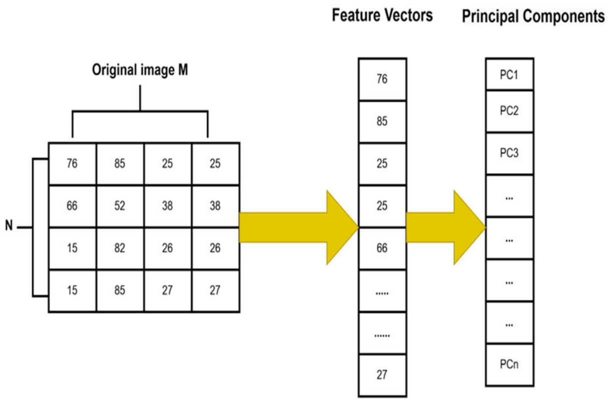

FIGURE 3. **Dimensional reduction with PCA.**

facial features like eyes, nose, and mouth in an Eigen feature layer. Mousavi et al. [41] used the nose tip as a reference point in the study and transformed the 3D face shape into a standard-sized image. It then used two-dimensional PCA (2D PCA) on this normalized image. The feature vectors representing the 3D facial shape were the eigenvectors corresponding to the largest eigenvalues. In the final step, an SVM classifier was used to recognise face images.

Eigenface: Eigenfaces is a popular method among appearance approaches for extracting facial feature points from face images [37]. This method utilises PCA, a widely acknowledged technique in face representation and recognition. PCA transforms a set of correlated variables into uncorrelated principal components or Eigenfaces, reducing the high dimensionality of data to a more manageable intrinsic dimensionality. This transformation aids in efficiently describing the data as illustrated in Figure 3. The Figure 4 showcases how facial features can be represented using a compact set of variables. PCA identifies the covariance matrix eigenvectors and projects the original data onto a lower-dimensional space defined by significant eigenvalues of eigenvectors. It is important to note that PCA is a well-established approach in the field and is commonly employed for FR purposes.

LDA: This is also known as "Fisher's Discriminant Analysis", is a popular technique in FR. Unlike PCA, which constructs a subspace to represent faces, LDA builds a subspace specifically to differentiate between the faces of different individuals [37]. LDA is frequently utilised for both dimensionality reduction and FR purposes [32]. PCA operates as an unsupervised technique, whereas LDA functions as a supervised learning method, utilizing available data information. LDA allows the assessment of crucial facial information to recognise human faces. The primary objective of LDA is to categorise face images into groups using features that most accurately describe them. Many LDA based FR systems face a challenge in terms of their optimality criteria which are often not directly linked to the ability of the system to classify the obtained feature representation [42]. The within-class scatter matrix ($\boldsymbol{S_w}$) is the mathematical construct used in LDA for feature extraction and classification. The within-class scatter matrix ($\boldsymbol{S_w}$) and the between-class scatter matrix ($\boldsymbol{S_B}$) for all samples across all classes are defined as follows:

$$\boldsymbol{S_B} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{c}} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^{\boldsymbol{T}} \tag{1}$$

$$\boldsymbol{S_w} = \sum_{\boldsymbol{x_k}}^{\boldsymbol{C}} \boldsymbol{M_i}(\boldsymbol{x_k} - \boldsymbol{\mu})(\boldsymbol{x_k} - \boldsymbol{\mu})^{\boldsymbol{T}} \tag{2}$$

where, μ denotes the mean vector of samples specific to class (i), $\boldsymbol{X_i}$ represents the set of samples associated to class (i), ($\boldsymbol{x_k}$) signifies the kth sample of that class, c is the number of distinct classes, $\boldsymbol{M_i}$ is the number of training samples in class I, $\boldsymbol{S_B}$ describes the scatter of features around the overall mean for all face classes, and $\boldsymbol{S_w}$ describes the scatter of features around the mean of each face class.

The objective is to maximize the ratio $\det|S_B|\det|S_w|$, which translates to minimizing S_w while maximizing S_B. This ratio is a key criterion in LDA to find the optimal feature representation for classification in FR systems.

ICA: Like PCA, ICA is a well-established subspace method widely employed in various fields. Like PCA, ICA also projects data from a high-dimensional space to a lower-dimensional form, making it a valuable technique for feature extraction. ICA is often considered a generalization of PCA and is primarily used to address challenges in signal processing [38]. ICA is regarded as a method applicable in FR tasks, where crucial information may reside in high-order relationships among pixels. PCA treats images as random variables following a Gaussian distribution and focuses on minimizing second-order statistics during the data transformation process. In the case of a non-Gaussian distribution, PCA will not match significant variances to its basis vectors. In contrast, ICA goes beyond PCA by addressing second-order dependencies and higher-order dependencies present in the input data. It aims to discover a basis where the data exhibit statistical dependence, making it effective for non-Gaussian distributions [43].

The ICA technique is utilized to compute the basic vectors of a given space. Its objective is to execute a linear transformation to minimize the statistical dependence among these vectors, enabling the analysis of independent components. ICA aims to ensure that the computed basic vectors are not orthogonal to each other. Additionally, when acquiring images from diverse sources as uncorrelated variables, ICA enhances efficiency. This is because ICA processes images as statistically independent variables, allowing for a more effective analysis. ICA outperforms PCA in several aspects. Unlike PCA, ICA is sensitive to higher-order data and doesn't focus solely on higher variance. It also produces a superior probabilistic model compared to PCA. Furthermore, the ICA algorithm is iterative in nature [43]. Despite of its advantages, ICA encounters challenges in handling large datasets. Additionally, it is reported to face difficulties in accurately ordering the source vectors.

2) NON-LINEAR TECHNIQUES

Non-linear techniques in face recognition involve sophisticated mathematical models capable of capturing intricate patterns and relationships within facial data, surpassing the limitations of linear methods. The Kernel PCA (KPCA) and Fisher's Linear Discriminant (FLD) were used in the literature as a non-linear technique in FR.

KPCA: In [44], the authors have proposed an enhanced version of PCA that utilizes kernel methods. Unlike PCA, which computes the covariance matrix, KPCA calculates the Eigenfaces or Eigenvectors from the kernel matrix, making it a powerful technique for nonlinear dimensionality reduction. Moreover, KPCA represents a transformation of the PCA technique into a high-dimensional feature space achieved through the associated kernel function mapping. Kernel Linear Discriminant Analysis (KDA) [45] is an extension of the linear LDA technique using kernel methods, similar to the kernel extension of PCA.

KDA applies kernel tricks to enhance the discriminative power of traditional LDA, making it effective for nonlinear classification tasks. The authors in [44] introduced the utilization of Discrete Cosine Transform (DCT) [46] in both global and local FR systems such as Gabor-KLDA [47], Wavelet transform (WT), radon transform (RT), and convolutional neural networks (CNN), Joint transform correlator-based two-layer neural network, Kernel Fisher discriminant analysis (KFD) and KPCA [44], Locally linear embedding (LLE) and LDA, Nonlinear locality preserving with deep networks, Nonlinear DCT and kernel discriminative common vector (KDCV) [48].

FLD: FLD was utilized to establish distinct and separable classes within a lower-dimensional space [49]. This approach, referred to as FisherFaces, has been proven to outperform the Eigenfaces method on datasets such as the Harvard and Yale Face Databases [50]. A significant hurdle faced in traditional LDA pertains to small sample size datasets [51]. To counter this issue, Wang and Tang introduced a solution called the dual-space Linear Discriminant Analysis approach [52]. W. Liu and colleagues [53] introduced the Singular Value decomposition(SVD) Updating based on Incremental PCA method for FR. A novel technique named the Diagonal PCA (DiaPCA) method has been introduced, aiming to derive optimal projective vectors directly from diagonal face imagery without the need for image-to-vector transformation.

The study explored the effectiveness of the ICA algorithm in FR, proposing a method that utilizes ICA as a feature extractor and SVM as a classifier for FR [53]. The research compared ICA, SVM and PCA methods on two distinct face databases, finding their accuracies to be comparable. However, the system employing Kernel PCA and an SVM classifier demonstrated a lower error rate when applied to the ORL database [54].

However, novel approaches have been proposed for facial recognition using PCA, Fisherfaces, the traditional LBP was proposed to overcome the limitations of the Kohonen approach. Particularly, LBP stands out due to its simple theory, computational simplicity, invariance concerning grayscale transformations, powerful rotation-invariant analysis, and excellent discrimination between various texture patterns [55]. However, LBP's robustness in face detection is challenged by factors such as ''noise, illumination variation, background, pose, scale, and occlusion'', making it less reliable than algorithms like VJ for face detection, as highlighted in references [56] and [57].

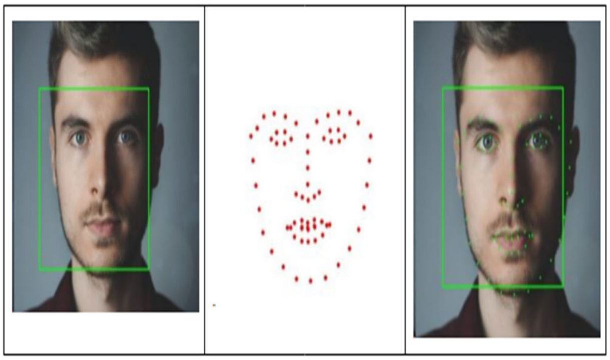

FIGURE 4. Landmark detection.

Table 3 provides a comprehensive overview, highlighting the distinctions among the latest methodologies used in both linear and non-linear approaches. This comparative analysis delves into the nuanced features and database, shedding light on their advantages, and limitations. The table serves as a valuable resource for gaining insights into the diverse landscape of contemporary techniques within the realms of linear and non-linear methodologies.

3) LANDMARK METHODS

Indeed, in FR, landmarks are crucial reference points used to identify and analyze facial characteristics. These specific points, such as corners of the eyes, tip of the nose, and corners of the mouth are anatomically significant and can be precisely located on a person's face. Landmarks serve as essential cues for facial recognition algorithms, aiding in accurately identifying and analyzing facial features.

In FR, landmark detection involves using algorithms to identify and mark specific points on a person's face in images or video frames. Once these landmarks are detected, they serve as reference points for extracting various facial features, such as the distance between the eyes or the angle of the mouth. The sample of retrieving landmark from the given image is shown in Figure 4. The extracted features are utilized to generate a distinct representation of an individual's face, commonly referred to as a facial template or facial signature. It is possible to create a 3D face feature descriptor by combining several measurements, including head width, nose height, nose width, nose depth, eye separation, and curvatures [58]. In the realm of 3D FR experiments, the distances between these feature descriptors of 24 faces are computed for detailed analysis [59], [60]. The Distinctive Landmark-

TABLE 3. Cutting-edge approaches in linear and non-linear methodologies.

Technique	Dataset	Method	Advantages	Limitations
LPQ and LDA[28]	MEPCO	SVM	Good accuracy	Computation time
PCA and Gabor filter [43]	FERET	Cosine metric	Pose	Precision
PCA[41]	YALE	SVM	Reduce the dimensionality	Recognition rate
KPCA and GDA[44]	UMIST face	SVM	Excellent performance	High error rate
CNN[34]	ORL	—	High recognition rate	Pose
Haar-cascade[20]	GreyScale Image	—	Object Localization	Recognition rate
PCA+LDA[48]	ORL YALE GB	Gabor filter	Reduce the dimensionality and recognition rate	Spoofing
LBP+ HOG[31]	Far-infrared	KNN	Feature Fusion	Accuracy
ICA[43]	ORL	SVM	Computational simplicity	Noise, Illumination
KPCA and KDA[45]	UMIST face	CNN	Complexity	Processing time

based Face Recognition (DLFR) is a specialized approach developed to handle the considerable challenge posed by the striking resemblance in facial appearances, especially in the case of twins, within FR tasks. The system incorporates distinctive features derived from a modified scale-invariant feature transform algorithm, focusing on the number of key points. Emphasizing the most unique landmark region of the face, this approach ensures a specialized and accurate representation for FR tasks. To optimize these features, a slightly modified genetic algorithm is utilized to determine their respective weights. Subsequently, the weighted features undergo processing through a SVM classifier, enhancing the precision and effectiveness of the recognition system [59]. In the literature the Point Distribution Model (PDM), plays a vital role in landmark based detection.

PDM: It is a shape description technique that heavily depends on landmark points. These landmark points are annotations made on specific locations of an image, aligning with corresponding locations on the shapes within the training set images. In the PDM, the shape of a face is constructed by placing landmark points on the facial features of the training images, forming a representation of the shape and structure of the face [60]. The model typically encompasses a global face shape, incorporating the formations of eyes, ears, nose, and other facial elements, as depicted in Figure 5.

PDM offers the flexibility to be fitted with various face shapes and results in a compact representation of the face. Nevertheless, constructing the training set by accurately marking the landmarks on facial features can be tedious and error-prone in many cases [61], [62].

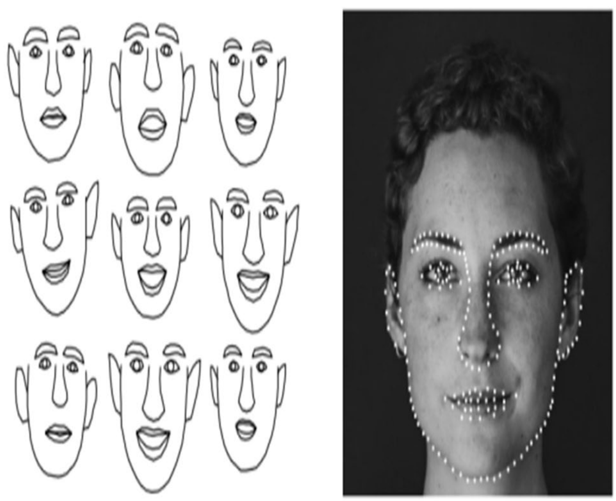

FIGURE 5. Global face shapes.

In order to highlight important facial regions, including landmarks, the study presents a novel self-attention distillation framework that aligns low-quality photos with their high-quality counterparts inside the feature space. This method efficiently regularizes the network in unconstrained contexts to learn a unified, quality-invariant feature representation [63].

4) HYBRID APPROACH

Hybrid approaches, which combine local and subspace features, harness the strengths of both techniques. Integrating local and subspace methods in these hybrid approaches can significantly enhance FR systems' performance, providing more robust and accurate results.

Fathima et al. [64] introduced a hybrid approach which combines Gabor wavelet and LDA for FR. In this method, grayscale face images are approximated and reduced in dimension, offering an innovative approach to enhance FR accuracy. The authors utilized a bank of Gabor filters with diverse orientations and scales to convolve the grayscale face image. Subsequently, a subspace technique known as 2D-LDA is applied to maximize inter-class space and minimize intra-class space, thereby enhancing the discriminative power of the extracted features for FR. For classifying and recognizing test face images, the authors employed the KNN classifier. In this method, the recognition task involves comparing the features of the test face image with those in the training set. Experimental results highlighted the robustness of this approach, particularly under varying lighting conditions.

Hu et al. [65] introduced a unique approach for FR centered around a fog computing-based scheme, specifically tailored to address FR challenges within the Internet of Things (IoT) context. In their method, a FR system generates a matrix of identities for an individual, paving the way for innovative solutions in the IoT domain. After the initial generation of the identity matrix, the proposed fog computing-based model determines individual identity. Experimental results showcase the model's bandwidth utilization efficiency and its remarkable accuracy, reaching up to 96.77%. This achievement represents a significant advancement compared to previous methods in the field of FR.

Yan et al. [66] introduced a robust FR technique named "Multi-Sub-Region-Based Correlation Filter Bank (MS-CFB)". This method emphasizes efficient feature extraction by independently extracting local features from different face sub-regions. Once local features are extracted from these sub-regions, they are concatenated to generate optimal overall correlation outputs. This innovative approach reduces complexity, higher recognition rates, and superior feature representation for recognition purposes. Comparative evaluations with several state-of-the-art techniques on various public face databases underscore the effectiveness of this method.

A collection of features known as biorthogonal wavelet entropy was presented by Zhang et al. [67] to produce a multiscale representation of facial structure. These characteristics improvise the analysis of facial features. Using a stratified cross-validation strategy, the authors trained a fuzzy multiclass support vector machine. This technique probably contributed to the robustness and classification accuracy of facial recognition system. MagFace, a class of loss functions designed to learn a universal feature embedding where the magnitude indicates the quality of the given face [68].

5) THE MODERN ERA OF FR USING DEEP LEARNING

In recent decades, DL has gained substantial attention in the realm of FR. Recent advances have led to the proposal of numerous methods based on DL techniques [69]. As depicted in the Figure 6, there has been a notable increase in the adoption of deep learning methods. Leveraging the significant utility of these methods, numerous prospective studies could be conducted in this field.

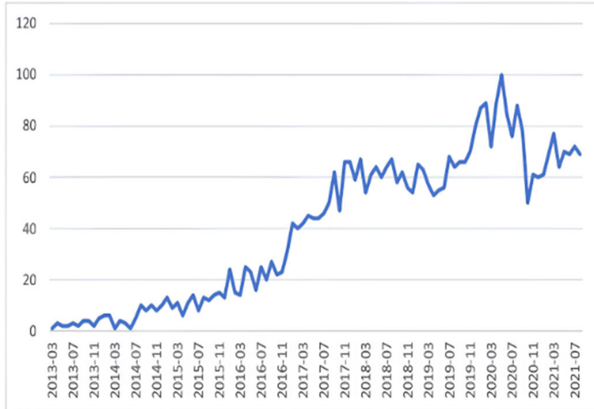

FIGURE 6. Growth of deep learning [109].

A CNN–LSTM–ELM hybrid deep architecture was presented by Sun et al. [70] and is intended for sequential Human Activity Recognition (HAR). Extreme Learning Machines (ELM), CNN, and LSTM networks are combined in this novel framework. Using the OPPORTUNITY dataset which comprises 46,495 training samples and 9,894 testing samples the researchers evaluated the CNN–LSTM–ELM structure. Each sample is represented as a sequence. The GPU used for the training and testing of the model had 1536 cores, a clock speed of 1050 MHz, and 8 GB of RAM [70].

The Multimodal Deep Face Representation (MM-DFR) framework, which uses of CNNs, was introduced by Ding et al. [71]. The original appearance face image, a rendered frontal face produced by a 3D face model (representing appearance and local facial features, respectively), and uniformly sampled image patches are among the several inputs that this method processes. The primary steps of MM-DFR framework is illustrated in Figure 7. In this study, a three-layer Stacked Auto-Encoder (SAE) technique was used. This SAE technique serves the purpose of compressing the high-dimensional deep features into a condensed face signature.

Facial landmarks and distance-based features are computed using deep learning techniques, which makes facial expression classification easier [72]. This solution allows developers to create flexible, multimodal, cross-platform machine learning pipelines by efficiently estimating 468 3D facial landmark points in real-time [73], [74].

3D face landmarks can be extracted from pictures and videos using a transfer learning technique. This procedure entails training a neural network created especially for this purpose [73]. Features are extracted using the distances between each facial landmark point and a selected reference landmark to categorize the emotions expressed on human

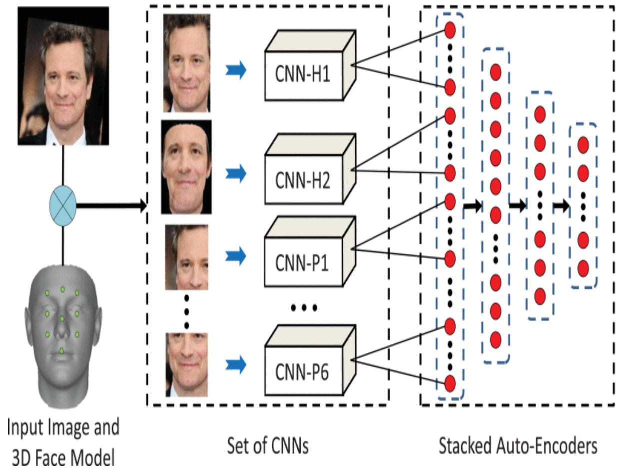

FIGURE 7. Schematic representation of MM-DFR approach [71].

faces. The reference landmark used in this computation for measuring distances is the landmark that corresponds to the tip of the nose as shown in Figure 5. The Euclidean distance of each landmark from the nose tip is computed as follows:

The reference landmark is set as the nose tip, and the Euclidean distance of the ith landmark from this reference point is computed using the following formula:

$$\mathit{Distance} = \sqrt{\left(x_i - x_{nose\ tip}\right)^2 + \left(y_i -, y_{nose\ tip}\right)^2 + \left(z_i - z_{nose\ tip}\right)^2}.$$

Here, $(x_i,\ y_i, z_i)$ represents the coordinates of the ith facial landmark, and $(x_{nose\ tip}, y_{nose\ tip}, z_{nose\ tip})$ represents the coordinates of the nose tip landmark.

The approach [75] combines statistical facial dynamics features obtained from the locations of facial landmarks during smile expressions with appearance-based features extracted using Deep Convolutional Neural Networks (DCNNs). This hybrid approach integrates both visual appearance cues and facial movement patterns, resulting in a more robust and accurate smile recognition system. To assess performance and robustness, three different DCNNs were evaluated under severe image distortions. The experimental results demonstrate that in challenging conditions, the accuracy of FR using solely DCNNs based features significantly decreases. However, integrating of facial dynamics features with DCNNs based features compensates for this decline in performance and substantially enhances accuracy [75].

The researchers introduced an innovative method known as the Deep Unified Model (DUM) to improve FR [76]. DUM combines CNN with edge computing techniques, enhancing the efficiency and accuracy of FR systems. The researchers trained their model using the widely-used LFW dataset and evaluated its performance in a student attendance system that employs FR technology. Their proposed technique demonstrated impressive performance, particularly for frontal face images.

Li et al. [77] introduced an innovative hybrid approach for person recognition in photo albums. Their method combines deep convolutional neural networks with carefully crafted features extracted from each individual's image. Li et al. [77] utilized multi-modality features obtained through a weighted average fusion of DCNNs and hand-crafted features. In their recognition process, these fused features were classified using a pre-trained SVM. While their experimental results demonstrated the method's effectiveness, it's worth noting that the approach, while successful, suffers from computational inefficiency due to the requirement of computing both hand-crafted and deep convolutional hybrid features for optimal performance.

In the Community project, a sophisticated deep learning model is employed, utilizing a hierarchical CNN architecture. This unique design allows the model to grasp intricate and detailed features intricately connected to each cohesive network within the community. Additionally, the research paper introduces two ground breaking techniques: the Face Co-occurrence Frequency algorithm, which quantifies the presence of individuals in images, and a distinctive photo ranking method. This method evaluates the strength of relationships between individuals in a predicted social network [78].

The study presented in [79] proposed a novel method known as Cluster-based Large Margin Local Embedding (CLMLE). This approach is designed to learn highly discriminative deep representations. CLMLE enforces a deep neural network to preserve inter-cluster margins both within individual classes and across different classes. This achievement is realized by introducing angular margins between cluster distributions on a hypersphere manifold. The experimental results illustrate that CLMLE, when coupled with a straightforward k-nearest cluster algorithm, leads to a substantial improvement in accuracy for FR and face attribute prediction tasks, especially those involving imbalanced class distributions [79].

The Deep Landmark Identification Network (DLIN) was presented by Gilani et al. [80]. It uses a binary classification loss to identify 11 facial landmarks. The training dataset is created artificially with the commercial program FaceGen and contains known landmark locations. This innovative approach aims to accurately identify facial landmarks through a specialized deep learning network and synthetic data generation. This dataset consists of 3D faces that have been augmented with diverse shapes to enrich the training process. More precisely, the dataset includes variations in age, masculinity/femininity, weight, height, as well as four distinct facial expressions (surprise, happiness, fear, and disgust) and five different poses (frontal, ±15° in pitch, and ±15° in roll). This comprehensive dataset allows for a robust training experience, incorporating a wide range of facial variations and expressions, enhancing the model's ability to handle diverse real-world scenarios. A wide array of factors is considered when crafting a diverse and all-encompassing training dataset for the DLIN [80]. Specifically, each 3D face generated undergoes a transformation into a spherical representation. This transformation derives three distinct image channels such as depth, azimuth, and elevation. These channels are subsequently utilized as input data for the DLIN [80]. This meticulous process ensures that the network is trained comprehensively, encompassing various facial features and expressions within its dataset, enhancing its accuracy and adaptability. The process involves segmenting each 3D face into five regions using geodesic level set curves, guided by detecting five key fiducial landmarks. This segmentation method ensures a structured division of the facial surface, enabling detailed analysis within the DLIN. Within each segmented region, distinctive key points are extracted, establishing dense correspondences across faces. These correspondences are then utilized to create a Region-based 3D Deformable Model (R3DM). The method's main goal in the context of 3D FR is to minimize the cosine distance between the parameters of the probe and gallery faces in the R3DM model. This meticulous process ensures a precise and discriminative comparison, leading to accurate FR results.

In order to improve face recognition in unconstrained situations, a novel technique termed bypass enhanced representation learning (BERL) is presented in this study [81]. This approach integrates two auxiliary bypasses, a blind inpainting bypass and a 3D reconstruction bypass, to enable both supervised and self-supervised learning. These workarounds aid robust feature learning for face recognition. Authors have developed a novel face detection and recognition model for drones to enhance face identification accuracy when query photographs are obtained from great heights or distances that do not reveal much facial information about persons [82].

In order to categorize individual interactions and body language from forceful postures, left and right sides, head orientation, and angular orientation of forms via webcam, the authors have suggested a real-time face matching technique based on the YOLO-V5 framework for image processing [83]. It takes different facial orientations and makes use of multi-pose human patterns. The paper addresses the challenge of unidentifiable face images within training datasets by leveraging the variation in recognizability based on image quality. Their approach involves two key strategies: first, they utilize feature norms as indicators of image quality, and second, they dynamically adjust the margin function based on these feature norms to control the gradient scale assigned to different attributes of images. Through comprehensive evaluations on datasets with varying image qualities, their adaptive loss mechanism demonstrates superior performance, achieving state-of-the-art results particularly on datasets containing mixed and low-quality face images [84].

The IDL-ERCFI technique used in the study [85] relied on intelligent DL methods. Its primary objective is to distinguishing and classifying ethnicity based on facial photos, showcasing the application of advanced DL techniques in the field of ethnic classification [85]. The IDL-ERCFI technique utilized the face landmarks for photo alignment before processing the images through the network. In this model, aneException network serves as the feature extractor. Given that the features extracted are high-dimensional, the paper employs PCA to reduce the feature dimensions. This approach effectively tackles the challenge posed by the high dimensionality of the dataset also ensuring a more manageable and streamlined set of features for further analysis and processing. The parameters of the model are fine-tuned using the Glow Worm Swarm Optimization (GSO) technique.

PCA and LDA utilize global information to generate a scatter matrix that transforms data from a high-dimensional space to a lower-dimensional subspace, maximizing the variance of the reconstructed data. However, the effectiveness of PCA and LDA declines significantly when the data is affected by noise and outliers. To address this problem, several improved versions [86], [87], [88], [89], and [90] have been developed to enhance the algorithmic robustness in face recognition. To fully utilize local information, various manifold learning-based algorithms were developed to achieve face recognition using the concept of topological manifolds. Locally Linear Embedding (LLE) [91] is a common non-linear manifold learning method known for effectively maintaining the original data's local geometric structure. Building on this idea, several extension methods have been proposed to enhance efficiency, such as Laplacian Eigenmaps (LE) [92] and Isomap [93].Despite achieving notable face recognition performance, these dimensionality reduction techniques still have certain limitations. To address these, He and Niyogi [94] introduced Locality Preserving Projections (LPP) for dimensionality reduction and feature extraction using a linear projection. Although LPP can preserve the locality information of the original data, the resulting linear projections are not orthogonal. To obtain robust and discriminative features for the face recognition challenge, a novel reweighted robust and discriminative latent subspace projection (ReRDLSP) framework is proposed [95]. This new structure seamlessly integrates the advantages of relaxed ridge regression, sparse representation, reweighted low-rank regularization, and latent subspace learning into a single framework. Additionally, Yang et al. [96] proposed an approach to orthogonal autoencoder regression. The success of nuclear norm-based matrix regression techniques has shown significant performance in image recognition.

Table 4 outlines the mix of techniques discussed in this section. These methods aim to make recognition systems better and more accurate. A comprehensive overview highlighting the distinctions among the latest methodologies used in deep learning approaches is given in Table 5.

In summary, Table 6 and Table 7 presents a comprehensive comparison of the reviewed algorithms and their respective attributes and trade-offs. From Table 6 and Table 7, it is inferred that, various algorithms address challenges in diverse ways to enhance accuracy and detection rates in FR. As per the current scenario, the available algorithms exhibit performance variations and strengths and weaknesses in face

TABLE 4. Comparison of state-of-the-art methods for hybrid approach.

Technique	Dataset	Method	Advantages	Limitations
Gabor wavelet[64, 100]	Yale, AR, CMU PIE, Yale B	Gabor Fisher classifier	Discrimination	Reliance on Data
CNNs and SAE[75]	LFW	—	Elevated accuracy	Intricacy
CNN–LSTM–ELM[70]	OPPORTUNITY	LSTM	Acquiring features	Extended processing duration
MS-CFB[66]	Yale	—	Simplified complexity	Pose
MM-DFR[71]	—	CNN	Condensing the deep features with high dimensionality.	Unconstrained
Fisher + SIFT[26, 119]	LFW	Mahalanobis matrix	Resilient	Singular feature category

TABLE 5. Comparison of state-of-the-art methods for deep learning approach.

Method	Dataset	Architecture	Accuracy
LBP[55]	CMU-PIE	VGGnet	97.65%
AdaBoost,SVM[56]	PolyU-HSFD	SI_CNN	88%
LBP,HOG, KNN[75]	IIITD kincet	AlexNet	86%
PCA,DCT[85]	CASIA-Webface and LWF	VGGnet	94.8%
Deep Unified Model [76]	LFW	CNN	—
DCNN[75]	Bosphorus	DCNN	99%
Transfer Learning[73]	LFW	Auto encoder	98%
Deep CNN [78]	CASIA	CNN	99.1%
PA-GAN [73]	IJB-A	–	99%

detection. Moreover, it is evidenced in the literature that, some studies grappled with overfitting issues while others excel in computational efficiency.

B. VIDEO-BASED FACE RECOGNITION

In contrast to traditional FR methods, video-based FR involves analyzing video data instead of static images. Like its conventional counterpart, video-based FR techniques aim to either identify individuals within a video (identification) or determine whether two subjects in different videos share the same identity (verification). The popularity of video-based FR has surged due to its diverse applications. Indeed, this field has attracted considerable attention in both computer vision and biometrics. Its applications extend to various domains including visual surveillance, access control, and video content analysis. Researchers and practitioners leverage these advancements to enhance security, automate processes, and gain valuable insights from visual data. With the proliferation of cameras in various locations worldwide and the widespread use of handheld devices capable of capturing videos, there is a continuous influx of vast video datasets. This accelerated the further research and advancements in this area.

Figure 8 provides one of the studies proposed for video-based FR [97]. In this study, the test video is taken from a

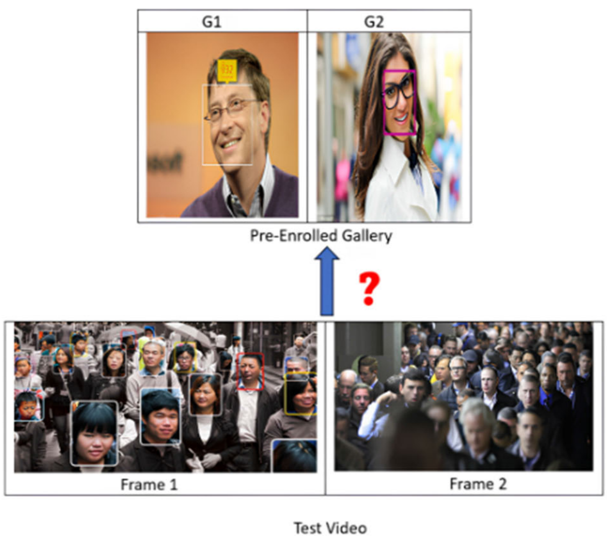

FIGURE 8. Pipeline of video based face recognition.

security camera, and the gallery is made up of static images which were submitted ahead of time. Especially in open-set scenarios, the objective is to identify and match each face in the video with a subject in the gallery or classify it as an unseen class. It is noteworthy to emphasize that this

TABLE 6. Comparative evaluation of face detection algorithms.

Methods	Advantages	Disadvantages
Artificial Neural Network	Capable of handling incomplete data	High computational cost
Decision Based Neural Network	Enhances comprehension of structural complexity	Limitation on face orientation
Fuzzy Neural Network	Higher accuracy	Dependent on linguistic rules
Eigenfaces	Straightforward and effective	Vulnerable to changes in image scale
Probabilistic Eigenspaces	Manages significantly increased occlusion levels	Effective primarily for rigid faces
Fisherfaces	Works well with images under diverse lighting conditions and facial expressions	Strongly influenced by input data
Tensorfaces	Efficiently maps images irrespective of variations in illuminations and facial expressions	Requires training with appropriately labelled multimodal training data
Principal Component Analysis	Excels in a controlled or limited environment	Affected by changes in scale
Support Vector Machine	Low risk of over-fitting	Performs inadequately with a dataset containing noise in the images
Discrete Cosine Transform	Computationally cost-effective	Depends on quantization
Locality Preserving Projection	Swift and well-suited for practical applications	Prone to interference from noise and outliers
Independent Component Analysis	Iterative	Encounters challenges in managing a large volume of data
CNN	Learning Feature	Limited resilience to variations in images

TABLE 7. Comparative evaluation of various approaches.

Approaches	Advantages	Disadvantages	Challenges Handled
Linear	These techniques demonstrate good performance when utilizing frontal views of faces.	Responsive to rotations and translations of face images.	Varied illumination conditions, scaling, and facial expressions.
	Recognition is both effective and straightforward.	Limited to classifying faces that are already present in the database.	
	Dimensionality reduction to represent global information.	Slow speed in face recognition due to a lengthy feature vector.	
Non-Linear	Reducing dimensionality	Recognition performance is contingent on the selected kernel.	Different illumination [70,69], poses [48], conditions, scaling, facial expressions.
	They are applicable to supervised classification problems.	Implementation is more challenging compared to local techniques.	
	Automatically detects features in this approach (CNN and RNN).	Unsatisfactory recognition rate.	
Hybrid	Offers faster systems and efficient recognition.	Implementation is more challenging.	Pose, illumination conditions, and facial expressions
		Complex and computationally costly.	
Deep Learning	Enhanced accuracy and robustness.	Spoofing, Privacy	Pose, Occlusion and Privacy

task is made more complex and difficult by the presence of low-quality faces in the video frames. In literature, there are two methods, viz., set-based methods and sequence-based methods employed for video-based FR. The details of these techniques are presented below.

1) SET-BASED METHODS

In the set-based approach of FR, the frames extracted from the video are treated as a collection of individual image samples, with no consideration for their temporal order. The set-based FR methods have been developed specifically to tackle the FR challenges posed by low-resolution images often acquired from surveillance cameras. Moreover, the set-based approaches are categorized into methods that employ fusion before matching and those that use fusion after matching. Fusion before matching entails combining features extracted from each face image prior to the recognition process [98]. Whereas the fusion after matching technique combines the recognition results obtained from each individual image. In this approach, FR is performed independently on each frame, and the outcomes are integrated or fused afterward to make a final decision. The authors in [99] utilized the Gabor Wavelet Network [100] to detect and segment specific facial regions like the mouth, eyes, and nose. Subsequently, feature extraction was conducted using the Karhunen-Loeve transform, and the features extracted from these regions in each frame were classified using the KNN algorithm. To tackle the issues arising by low-resolution, low-contrast, and non-frontal face surveillance images/videos, researchers have proposed methods involving 3D face modelling. In [101], a 3D face model was created from multiple non-frontal frames in a video. Subsequently, person recognition was conducted utilizing a commercial 2D FR system. The experimental outcomes revealed a notable enhancement, indicating a 40% increase in the match ratio with the incorporation of 3D models. This highlights the effectiveness of 3D modelling in improving person recognition accuracy. This approach proved its effectiveness in enhancing recognition accuracy despite the difficulties posed by non-ideal imaging conditions.

In [102], a novel approach called Manifold-to-Manifold Distance (MMD) was introduced. This method represents sets of faces as manifolds or subspaces and models image sets using their second-order statistics for the purpose of image set classification. By employing this innovative technique, the research aimed to enhance the understanding and classification of sets of facial images, thereby contributing to advancements in the field of FR. Zheng et al. [103] introduced a novel technique called quality-aware principal angle-based subspace-to-subspace similarity. This method involves learning subspaces, enabling a more unique and precise measurement of similarity between different subspaces. Chen et al. [104] introduced a set-based algorithm that leverages sparse representation and dictionary learning techniques. Additionally, adaptive pooling method [105] also proposed, which aimed at flattening face sets. The techniques [103], [106] harness the contextual information within videos by propagating identity-related data based on graphical modeling.

2) SEQUENCE BASED METHOD

Sequence based face recognition method utilizes sequence of videos containing facial characteristics is used for face recognition in videos. In [106], the experiments which utilize the ENTERFACE dataset, which consists of 1300 videos featuring 44 individuals was presented. These videos capture short sentences expressing the six basic emotions. The researchers measured distances between characteristic points on the face, deriving 14 distinct attributes from these measurements. This detailed analysis of facial expressions provides valuable data for understanding and categorizing human emotions. These attributes were then averaged over the entire video and normalized. These normalized averages were employed for classification using Gaussian Mixture Models (GMM). In the context of the ENTERFACE database, the 1-best recognition accuracy was found to be 16 times better than random, indicating a substantial improvement. Even in the worst-case scenario, the system demonstrated performance 7 times superior to random.

Gong et al. [107] introduced a recurrent network for aggregating sequence features in face matching. Meanwhile, Zheng et al. [103] proposed a hybrid dictionary learning method that captures temporal correlations in video face sequences using dynamical dictionaries. Haque et al. [108] introduced a biometric person recognition system based on features derived from a pain expression model. Their study utilized a pain database comprising face videos of individuals experiencing shoulder pain during active and passive motion tests. Feature extraction was performed using the Facial Action Coding System (FACS), and classification was carried out using ANN [109]. The authors developed the Additive Angular Margin Loss (ArcFace) to produce highly discriminative features for facial recognition [110]. ArcFace provides a geometric explanation that makes sense because it lines up exactly with a hypersphere's geodesic distance. The ArcFace is evaluated using ten benchmarks, including a new large-scale image database with trillions of pairs and an extensive video dataset, and compared it to the most recent state-of-the-art facial recognition techniques. It was concluded that ArcFace continuously outperformed the state of the art with little computing overhead and simple to implement. Table 8 provides a concise overview of the set-based methods and sequence-based methods with respect to the type of facial features employed, classification methods, databases utilized, recognition rates, and key findings.

IV. EVALUATION METRICS, DATAETS AND TOOLS

This section details, the evaluation metrics, datasets, and tools used to assess the FR systems. In literature, various evaluation metrics have been proposed to measure the effectiveness and accuracy of FR systems. The prominent measures are listed below [2], [21]:

False Match Rate (FMR): This is also referred as False Accept Rate (FAR) which indicates the proportion of impostor (intruder) samples that are inaccurately identified as the legitimate identity in a FR system. FMR is a crucial metric for evaluating the security and accuracy of such systems, especially in scenarios where unauthorized access needs to be minimized. Lower FMR indicates a higher level of security as it means fewer impostor samples are being accepted as genuine

TABLE 8. Comparison of state-of-the-art methods for video based face recognition.

Feature extraction	Classification	Dataset	Accuracy %	Description
CNN[6]	SFDL, D-SFDL	YouTube Celeb DB	79	Observations indicate that the D-SFDL method is more successful than the SFDL method.
3D dynamic features[109]	Distance Measurement	ORL Database	—	The proposed method can potentially enhance the accuracy of face recognition.
FACS[111]	Artificial Neural Network	Painful data	87.9	The frames of individuals who experienced suffering were not consistently visually distinctive from those who did not experience suffering.
CNN features and geometric features[72]	SVM	Own database	96.2	The experimental results demonstrated that the transition frames outperformed the peak emotion frames in face recognition.
3D Morphable Model (3DMM)[118]	Distance Measurement	Survilance video	92.8	Improvement in Accuracy
Super resolution reconstruction , Eigenface[73]	Euclidean Distance	CMU	79	Super-resolution reconstruction yields superior face recognition.
CNN[19]	Fully Connected Layer	YTF	96.2	An attention-aware deep reinforcement learning approach is employed to discard frames that are not deemed useful.

False Non-Match Rate (FNMR): Which is also referred as False Reject Rate (FRR) that represents the percentage of genuine samples that are incorrectly rejected by a FR system. FNMR is a crucial metric as it measures the system's failure to recognize valid users which indicates the likelihood of genuine users being denied access. Lower FNMR values signify better performance ensuring that authentic individuals are correctly identified and granted access thereby minimizing the risk of false rejections.

Accuracy: It is a fundamental metric in evaluating the performance of any classification systems. It represents the percentage of samples that are correctly classified out of the total samples tested. A higher accuracy rate indicates a more reliable and precise system as it correctly identifies a larger portion of the samples. In the context of FR system, accuracy is a crucial measure as it directly reflects the ability of the system to correctly identify and authenticate individuals.

Genuine Accept Rate (GAR): Also known as True Acceptance Rate (TAR) which represents the percentage of genuine samples that are correctly accepted by a FR system. In other words, GAR or TAR is the complement of FNMR, calculated as (1 - FNMR). GAR/TAR is a critical measure in biometric systems, indicating the system's ability to accurately recognize and authenticate legitimate users, thus reflecting the system's reliability in accepting genuine samples.

Equal Error Rate (EER): It is a significant metric used in biometric systems. It refers to the point where the FMR and the FNMR are equal, meaning that the system's performance in falsely accepting and falsely rejecting individuals is balanced. At the EER, the system achieves an optimal trade-off between accepting genuine users and rejecting impostors, making it a crucial point for evaluating the overall effectiveness of a biometric authentication system.

Receiver Operating Characteristic (ROC): A curve representing the trade-off between the FRR and the FAR at different threshold values in a binary classification system. It illustrates how the performance of the system varies as the decision threshold changes. In the context of biometrics, the ROC curve can also be obtained by plotting the True Acceptance Rate (TAR) or GAR against the FAR. This curve provides valuable insights into the system's ability to distinguish between genuine users and impostors across various threshold settings. The ROC curve is a widely used tool in evaluating the performance of biometric systems, helping researchers and practitioners choose an appropriate operating point based on their specific requirements and priorities regarding false acceptance and false rejection rates.

Area Under the Receiver Operating Characteristic curve(AUROC): AUROC quantifies the ability of the system to discriminate between positive and negative classes across different threshold values. AUROC values range from 0.5 to 1.0. A value of 0.5 indicates random selection (no discrimination ability), while a value of 1.0 represents perfect classification, where the system can perfectly distinguish between positive and negative instances. The closer the AUROC value is to 1.0, the better the system's discrimination ability. AUROC provides a comprehensive summary of a system's performance, considering its ability to balance true positive rate (sensitivity) and false positive rate across various decision thresholds. It is a valuable metric for comparing and selecting different models or algorithms in various fields, including biometric authentication systems.

A. DATASETS

This section elaborates the publicly available datasets used in the domain of FR systems. It is seen the literature that, a variety of datasets have been employed to evaluate the performance of FR systems. The description of some datasets are discussed below:

CASIA WebFace is a dataset containing approximately 500,000 images of 10,000 subjects. Initially gathered by the CASIA group and later refined manually. This dataset follows a common pattern seen in collections based on celebrities or well known individuals. It exhibits a long tail distribution which indicates a few popular subjects have the majority of images while others are represented by only a few pictures.

VGGFace, introduced by the Oxford group which is designed to train deep learning models. It includes approximately 2.6 million faces of 2,622 individuals. Unlike CASIA, VGGFace has a balanced distribution in which each subject is represented by a thousands of samples that consists of high-quality frontal faces sourced from web engines. However, despite cleaning efforts this dataset suffers from some noise-related issues viz., outliers which leads to intra-class variance.

UMDFaces, introduced by Bansal et al. [19], utilized a combination of human annotators through Amazon Mechanical Turk (AMT) and pre-trained deep-based face analysis tools. This unique approach aimed to create medium-sized sets that were more challenging than existing ones like CASIA [113] and VGGFace [26]. Though UMDFaces dataset comprises both high-quality still images and video frames, still it susceptible to motion blur. This dataset annotates facial key points, face pose angles, and gender information. Specifically, it includes 367,888 face annotations in still images representing 8,277 subjects, along with 3.7 million annotated video frames extracted from approximately 22,000 videos which featured 3,100 subjects.

MS-Celeb-1M, initially introduced in the multimedia community and later adopted by the computer vision community [114]. It comprised of approximately 10 million images featuring 100,000 celebrities. Each celebrity in the dataset is represented by 100 images obtained from the Bing search engine using the celebrity's name as the query. The images were retrieved without any filtering applied to the search results. However, despite its vast size, MS-Celeb-1M faces significant challenges. It is important to note that MS-Celeb-1M was released specifically to learn from noisy labels and was never manipulated or organized for quality.

VGGFace2, an enhanced version of VGGFace, was developed to overcome the limitations of its earlier version [115]. This dataset contains 3.31 million images featuring 9,131 subjects including celebrities as well as notable individuals. In comparison to the original VGGFace, the average number of images per individual has slightly decreased to an average of 362.6 images. However, VGGFace2 is intentionally designed to encompass a wide range of poses, ages, and ethnicities while striving to minimize label noise [2], [3]. This reduction in label noise was achieved through a combination of manual and automatic processes.

IMDb-Face, a recently introduced dataset derived by meticulously cleaning label noise from MS-Celeb-1M and MegaFace [19], [114], [115]. This innovative dataset stands out as the largest noise controlled face collection currently available.

The FERET Dataset, is a pivotal dataset which was executed through a collaboration between the DARPA and the NIST. The FERET was established to foster algorithm development and evaluation within the field. Initially, FERET needed a standardized database of facial images for developing and testing evaluation methods. In 2003, DARPA released a high-resolution 24-bit colour version of these images to enhance the utility for facial recognition research and development. The database comprises 2,413 still face images which represents 856 individuals [113], [114].

Labeled Faces in the Wild (LFW), serves as a comprehensive dataset of face photographs specifically designed to address the challenges posed by unconstrained FR [19], [113]. This dataset comprises a vast collection of over 13,000 facial images sourced from the web with each face meticulously labelled with the corresponding name of the person. In this dataset, approximately 1,680 individuals are represented by multiple distinct photos. However, it is important to note that these facial images were initially detected using the Viola Jones face detector [7], [56], [57]. The dataset is organized into four sets of LFW images: one original set and three sets of aligned images. Notably, LFW-a and the deep funneled images consistently demonstrate superior performance across a wide range of face verification algorithms in comparison to the original and funneled images. This characteristic makes them invaluable for research and testing purposes.

The BU-3DFE dataset, comprises 2,500 3D facial scans belonging to 100 subjects including 44 males and 56 females. These subjects include diverse ages and ethnic/racial backgrounds. Each individual in the dataset is represented by one scan displaying a neutral expression as well as six basic non-neutral facial expressions, such as, surprise, anger, disgust, fear, happiness, and sadness. Additionally, each expression is captured at four different intensity levels providing a comprehensive dataset for facial expression analysis [15].

The YouTube Face Database (YTF) contains face videos specifically created for unconstrained FR. This database includes clips ranging from 48 frames to 6,070 frames in length with an average clip duration of 181.3 frames [3], [15], [109]. Each video was sourced from YouTube and there are an average of 2.15 videos available for different subjects. YTF serves as a valuable resource for studying FR algorithms in real-world uncontrolled settings.

The Yale Face Dataset offers a diverse collection of facial images for research purposes. It includes 165 grayscale images in GIF format featuring 15 individuals. The images are sorted into 11 categories featuring distinct facial expressions or configurations, such as, center-light with glasses, center-light without glasses, happy, sad, sleepy, normal, surprised, wink, left-light, and right-light. The database is available in two volumes: Yale Face A and Extended Yale Face Database B [15].

Yale Face A comprises 15 subjects (14 males and 1 female) exhibiting various facial conditions, including different expressions, viz., sad, normal, and happy. Additionally, it encompasses variations in lighting conditions such as left, right, or center light, and includes images with and without glasses.

Extended Yale Face Database B is a dataset containing 2,414 images of 38 subjects. These images do not have variations in Expression or Occlusion but are focused on extracting features suitable for illumination analysis. They are available in cropped versions, which facilitates the detailed research in illumination-related studies. Moreover, the Yale A database [109], [112] includes 165 images of 15 individuals, whereas, AR database features 2,600 images of 120 persons. The Yale B database encompasses nine distinct postures captured under 64 different lighting conditions. This extensive dataset is further subdivided into five subsets based on the angle between the light direction and the camera axis making it a versatile and comprehensive resource for various research applications.

The Gavab dataset is a comprehensive collection featuring 549 3D facial scans, capturing the faces of 61 adult European American subjects (45 males and 16 females) in remarkable detail. These high-resolution scans, obtained using a Minolta Vivid scanner, offer an impressive level of granularity at a resolution of 1.5mm per image [116]. The dataset covers a wide array of conditions including Pose, Occlusion, and Expression. For each subject, the dataset includes two frontal facial scans displaying a neutral expression. Additionally, there are four scans with a neutral facial expression, but the face is rotated into different postures. Furthermore, the dataset comprises three frontal non-neutral facial expressions. This diverse set of scans provides ample opportunities for in-depth analysis and research within the realms of facial recognition and expression analysis.

TinyFace is an FR benchmark created to facilitate large-scale research on Low-Resolution Face Recognition (LRFR), especially in deep learning frameworks. The 169,403 native low-resolution face photos, with an average size of 20 by 16, and 5,139 tagged facial identities are included in the TinyFace dataset, which is designed for 1:N recognition testing. These low-resolution photos were taken under a variety of lighting, occlusion, backdrop, and posing scenarios.

The IARPA Janus Benchmark-C (IJB-C) This dataset comprises of creative commons licensed face images and videos for 3,531 subjects, expanded by 1,661 new subjects. This dataset is also used for video-based face recognition and builds on the IJB-A dataset. It includes approximately 138,000 face images, 11,000 videos, and 10,000 non-face images.

The IJB-B dataset is a template-based face recognition dataset featuring 1,845 subjects with 11,754 images, 55,025 frames, and 7,011 videos. Each template comprises still images and video frames sourced from various origins. These images and videos, gathered from the Internet, are entirely unconstrained, exhibiting significant variations in pose, lighting, and image quality. Moreover, the dataset includes protocols for 1:1 template-based face verification, 1:N template-based open-set face identification, and 1:N open-set video face identification [103].

The IJB-S dataset is an open-source IARPA Janus Surveillance Video Benchmark accompanied by relevant protocols. It consists of images and surveillance videos of 202 subjects, collected from Department of Defense (DoD) training facility. The surveillance videos were recorded across multiple scenarios that reflect various real-world surveillance situations, particularly relevant to law enforcement and national security. The dataset includes over 10 million annotations in total [117].

The characteristics of different dataset has been analysed and tabulated in Table 9. Whereas, the abbreviations are interpreted as follows: V (various), FE (number of different facial expressions), IL (illuminations), PO (head poses), OC (occlusions, e.g. hand, hair, eyeglasses, beard...), TI (recording times), AC (accessories), BG (backgrounds), ET (ethnicities). Moreover, Depth indicates the number of images for each subject is high and breadth shows the number of subjects is high with as many images as possible for each subject.

In the pursuit of advancing unconstrained video-based FR systems, two datasets were recently proposed namely IJB-B [117] and the IJB-S. These datasets present a significant leap in difficulty compared to other datasets such as Multiple Biometric Grand Challenge (MBGC) and the Face and Ocular Challenge Series. Unlike the relatively controlled conditions of the latter datasets, IJB-B and IJB-S are captured in unconstrained settings, offering a more complex environment. These datasets feature faces exhibiting extensive intra/inter-class variations in aspects [103]. The challenges posed by these datasets push the boundaries of FR research, enabling the development of more robust and accurate video-based FR systems suitable for real-world applications, especially in surveillance contexts.

B. TOOLS

This section elaborates the various tools used in FR Techniques. The table 10 give the comprehensive analysis of the pivotal role of various tools to achieve accurate and reliable results by representing the forefront of FR technology, facilitating advancements in biometric authentication. The detailed analysis of significant tools used in various studies are given in Table 10.

V. PERFORMANCE ANALYSIS, FINDINGS, AND CHALLENGES

The analysis of various methods for face recognition under various unconstrained environments viz expression, illumination, pose and occlusion are presented in this section. Within the category of expression invariant-based methods, both local and global features are taken into account.

TABLE 9. Face dataset.

No	Name	Year	Source	Subject	Images	Size	Properties	Features	Format
1	CASIA WebFace	2014	CASIA group	10,000	500,000	V	Celebrities.	Long tail distribution	Image Files (JPG/PNG)
2	VGGFace	2015	Oxford group	2,622	2,600,000	V	Celebrities, face annotations with bounding boxes and pose.	High-quality frontal faces	Image Files (JPG/PNG)
3	UMDFaces	2017	Minolta Vivid 900	367,888 (still images) + 3.7M (video frames)	367,888 (still images) + 3.7M (video frames)	V	Breadth; video.	Mix of still images and video frames.	Image and Video Files
4	MS-Celeb-1M	2016	Multimedia Community	100,000	10,000,000	V	Breadth; celebrities; knowledge base.	Label noise and quality issues.	Image Files (JPG/PNG)
5	VGGFace2	2017	Oxford group	9,131	3,310,000	V	Depth; PO, BG, age, ET; celebrities.	Reduced label noise	Image Files (JPG/PNG)
6	IMDb-Face	2018	MS-Celeb-1M and MegaFace	59000	1.7M	V	Pose, age	Focuses on noise control and dataset quality.	Image Files (JPG/PNG)
7	AR Database	1998	(CVC), University of Alabama at Birmingham	126	4000	576 × 768,	FE: 4, IL: 4, OC: 2, TI: 2.	Images of each individual were captured on two separate days, with a 14-day interval between sessions.	Image Files (JPG/PNG)
8	FERET Database	2003	DARPA and NIST	856	2,413	V	high-resolution, 24-bit color version	Used for algorithm development and evaluation	Image Files (JPG/PNG)
9	LFW	2007	—	5.749	13.233	150 × 150,	Un-posed photos, mainly frontal views.	Unconstrained FR dataset with labeled images.	Image Files (JPG/PNG)
10	YouTube Face Database	2008	YouTube	–	47, 1910	V	All videos are encoded in MPEG4 at 25 fps rate.	Unconstrained FR dataset with face videos.	Video Files (MP4/AVI)
11	Gavab	2004	—	61	549	V	smile, laugh, and arbitrary expression	3D facial scans with various facial expressions, poses, and occlusions.	3D Scans (OBJ/STL)
12	FaceScape	2020	—	938	18760	V	age and gender	multi-view images	Image Files (JPG/PNG)
13	CPLFW	1018	crowdsourcing	3.968	11,652	250 × 250	PO, Celebrities.	—-	Image Files (JPG/PNG)
14	Mega Face	2016	Flickr	672.052,	4.7M	V	Breadth; the whole long tail; commonality.	—-	Image Files (JPG/PNG)
15	BU-4DFE	2008	NIST	106	606	1040 × 1329	FE	—-	Image Files (JPG/PNG)

Local features focus on rigid regions less affected by expression variations while global features include methods using morphable models to synthesize virtual faces with desired expressions and establishing unified expression models to transform target faces into neutral expressions. The pose invariant-based methods in the second category correct poses using rigid landmarks or synthetic faces. The third category, occlusion invariant-based, focuses on detecting and restoring occluded facial regions by extracting facial curves and non-occluded areas [118]. In this survey, Table 11 and Table 12 provide summaries of 3D FR under unconstrained conditions. Table 13 presents a comparison of the effectiveness of different methods in facial recognition across multiple datasets.

Table 14 compares Image-Based and Video-Based Facial Recognition Methods which offers insights into their key characteristics, advantages, and challenges. Image-Based Methods focus on feature extraction from static images which

TABLE 10. Role of tools for FR.

Tool Name	Description	Features
OpenCV	Open Source Computer Vision Library	Haar cascades for object detection LBP for face recognition DNN module for advanced face detection
Dlib	C++ Library with Python bindings	HOG for face encoding and face landmark detection
TensorFlow	Deep Learning Framework developed by Google	TensorFlow Object Detection API for face detection TensorFlow Lite is a streamlined version for deploying models on mobile and edge devices
PyTorch	Deep Learning Framework developed by Facebook	TorchVision library for face detection and recognition PyTorch Mobile for deploying models on mobile devices
MTCNN	Multi-task Cascaded Convolutional Networks	Cascade of CNNs for face detection, landmark detection, and alignment
RetinaFace	Real-time Face Detector optimized for mobile devices	Lightweight architecture for real-time processing Accurate face detection across various scales
SSD (Single Shot Multibox Detector)	Object detection framework for real-time processing	Single neural network predicts bounding boxes and class probabilities simultaneously
BlazeFace	Lightweight face detection model optimized for mobile and edge devices	Compact architecture for efficient inference Fast and accurate face detection

TABLE 11. Illumination and expression invariant methods.

Illumination	Expression
Utilize 3DMM to create multiple perspectives and diverse lighting scenarios.	Identify and locate 7 key points in the vicinity of the nose area.
Calculate the coefficients for spherical harmonic lighting and 3DMM.	Generate virtual faces with varying poses and orientations.
Leverage Relight GAN to perform instance-level face illumination transfer, ensuring realistic and accurate lighting adjustments for individual faces.	Enhance Realism in Facial Pose and Expression Adjustment
Apply advanced techniques such as dense landmark detection and semantic parsing to achieve realistic transformations in face illumination, enabling accurate and natural adjustments during illumination transfer and swapping processes.	Representation-Learning Wasserstein-GAN (RL-GAN) and Deep Transfer Network (DTN)

provides lower computational complexity and suitability for still or posed subjects. However, those techniques may face limitations in handling variations in lighting and pose.

On the other hand, video-based methods leverage information from video frames which results in increased robustness especially in dynamic scenes. These methods utilize temporal analysis and face tracking across multiple frames that offers improved accuracy for subjects in motion. However, those techniques also come with higher computational requirements and challenges related to efficient video processing and handling prolonged periods of occlusion. The Table 14 serves as a comprehensive guide for system designers and researchers in the field of facial recognition, aiding in selecting the most appropriate approach based on specific application requirements.

A. OPEN PROBLEMS

Unconstrained video-based FR remains a complex and unsolved challenge compared to still image-based FR. Several factors contribute to this increased difficulty:

1. In video-based FR, the testing data consists of videos which comprised thousands of frames, and every frame could

TABLE 12. Pose and Occlusion invariant methods.

Pose	Occlusion
Composite 3D Face Representation	Enhancing radial strings
Integrate 2D facial landmarks and a standard 3D face model to determine the orientation of the head, a process commonly referred to as head pose estimation.	Implement a thresholding method on RGB-D images
Pose-Weighted GAN	Use the block-based method for occlusion detection
Leverage GANs to create synthetic facial images featuring diverse poses.	Gappy Wavelet Neural Network (GWNN)
Utilize local and global adversarial DCNNs to comprehensively learn the UV map.	Image distortion analysis feature vector
Implement a system where two U-Net architectures are stacked to comprehensively learn the entire UV map.	Utilize a two-branch DCNN model to encode both texture and shape cues.
Multi-channel pose-aware CNN	MLFP database using latex and paper masks latex
Pose-Aware Models (PAM)	Custom-made flexible silicone masks were used to access the CS-MAD database.
—	Analyze heartbeat signal via rPPG
—	Extract the global rPPG from the facial region
—	Implement into practice a global noise-aware template learning multi-channel rPPG system.
—	Weighted Spatial–Temporal Representation

TABLE 13. Performance comparison.

Research work	Category	Method	Advantage	Database	Accuracy in %
Deng et al. [119]	SIFT keypoints	Riemannian kernel sparse coding	E, OC	'FRGC v2'	97.3
Abbad et al. [120]	Curve	Expression, time consumption	OC, missing data	'GavabDB'	99.18
Shi et al. [121]	Region (LBP-based)	LBP, SVM	P, OC	'Texas-3D'	96.83
Gilani et al. [80]	—	PCA	OC	'Bosphorus'	98.1
Liang et al. [122]	—	HK classification	P	'Bosphorus'	94.79
Papadopoulos et al. [123]	—	Face-GCN	—	'Face-GCN'	88.45
Haamer et al. [19]	CNN features and geometric features	SVM	P	'Own database (630 videos from 61 people)'	96.2
Rabah et al. [32]	Feature Extraction	PCA+LDA	OC	'JAFFE, AT&T, Yale, Georgia Tech, CASIA, Extended Yale, Essex'	99.25
Chu et al. [37]	—	C-RSDA	LR	'LFW database, SCface database'	—

potentially contain multiple faces. This situation poses a scalability challenge for video-based FR [103]. Creating an effective FR system requires fast performance from every component which often slows down the FR process and serves as a bottleneck in the pipeline.

2. Faces captured in unconstrained videos exhibit substantial variations in Pose, Occluson, Expression, and Illumination compared to still images. Therefore, the representations used for video faces must be robust enough to accommodate these diverse and significant variations [106].

3. Unconstrained videos often involve challenges like frequent occlusions, varying poses, and changing scenes, making it exceptionally difficult to associate faces accurately. If a set of associated faces contains multiple identities, the performance of FR systems is adversely affected [24].

TABLE 14. Comparison of image-based and video-based facial recognition methods.

Method Type	Key Characteristics	Advantages	Challenges
Image-Based	1. Feature extraction from static images 2. Single frame analysis 3. Face detection and feature matching	1. Less computational complexity 2. Suitable for applications with still or posed subjects	1. Susceptible to variations in lighting 2. Limited in handling pose variations 3. Limited effectiveness in real-time
Video-Based	1. Utilizes information from video frames 2. Temporal analysis for dynamic features 3. Face tracking across multiple frames	1. Increased robustness in handling 2. Temporal variations and dynamic scenes 3. Improved accuracy in recognizing subjects in motion	1. Higher computational requirements 2. Need for efficient video processing 3. May face challenges in handling occlusion and Expressions

4. In videos, each identity may have a different number of faces, resulting in varying-sized face sets or sequences after association. The challenge arises during face matching when it's difficult to encode these sets or sequences, which have different lengths into a consistent and fixed-size representation [108].

B. CHALLENGES

Though FR is extensively applied in real-time automation and security applications still it suffers from various challenges in terms of human face, an intricate entity, facial hair, eyewear, lighting conditions, and muscle deformation which undergoes constant short and long-term changes. From the existing literature it is identified that extreme angles significantly decreased system accuracy. Even slight changes in head pose, including roll, pitch, and yaw, can drastically affect FR accuracy.

Illumination variations, often called lighting effects, occur due to different environments impacting facial recognition systems, like day or night. Changes in illumination can create additional light or dark patches in the region of interest, significantly affecting the accuracy of facial recognition. variations in lighting conditions led to a notable decrease in FR accuracy, showcasing the challenge posed by illumination variations.

Aging conditions impact facial features due to skin, tissue, and muscle changes. As faces naturally change over time, identifying individuals in face images under long-term aging effects poses a significant challenge. Occlusion, often caused by eyeglasses, sunglasses, hats, scarves, and other objects covering the face, is a common challenge in real-world FR scenarios. Handling very large-scale systems is daunting, especially when databases contain hundreds of millions or over one billion face images [124]. Ensuring the efficiency of an FR system with such massive real-time databases poses a significant challenge.

The privacy issues and the ethical implications of FR technology becoming more and more important, it is critical to perform a thorough analysis in order to identify potential problems and suggest workable solutions [125]. Because FR technology has the potential to be used for widespread surveillance and invasion of personal privacy, it has raised serious concerns about privacy breaches and moral quandaries. The increasing use of FR systems in public areas, workplaces, and consumer apps has sparked debates about consent, ethics, and data security. Data security and privacy protections are also sparked by the possibility of data breaches and illegal access to private biometric information. To overcome these worries, strict privacy regulations and moral standards must be incorporated into the creation and use of FR technology.

VI. CONCLUSION AND FUTURE WORK

Face recognition presents a complex challenge in the field of computer vision, attracting significant attention due to its wide-ranging applications across various domains. This survey delves into the extensive literature on FR, examining the essential stages of the FR process and highlighting numerous challenges that profoundly influence system performance. The survey offers a comprehensive overview of FR systems, tracing their development throughout history. Moreover, a taxonomy of facial recognition methods and summarizes popular facial datasets used for training and testing these systems. The survey meticulously examines image-based and video-based FR methods, providing detailed comparisons of various approaches.

Future Directions

The following key points are identified as future research challenges in FR domain:

1. While significant attention has been given to handling facial expressions in face recognition, the challenges posed by variations in pose and occlusions remain underexplored and require further research.

2. Achieving cross-resolution, cross-age, and cross-sensor 3D face recognition remains a significant challenge. It is common for probe 3D faces and gallery 3D faces to be acquired with different sensors at different times, resulting in varying resolutions and noise levels. Developing methods to achieve accurate, robust, and efficient 3D face recognition under these conditions is still an unsolved problem.

3. Diverse modalities and descriptors should be incorporated to enhance the existing system and bolster its resilience against attacks.

4. Fusing generative AI with 3D facial detection can lead to better development of FR system. Numerous applications

have already been introduced in the field of 3D model generation, demonstrating the versatility of generative AI.

5. Robust privacy controls and ethical requirements are anticipated to be used in the development and implementation of FR technology.

ACKNOWLEDGMENT

The authors have no conflict of interest to declare.

Funding: we have not received any funding.

REFERENCES

[1] M. J. Bhaskar, Y. Venkatesh, R. S. B. Pranai, and M. Rohith, "Face recognition for attendance management," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 4, pp. 964–968, Apr. 2020.

[2] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020.

[3] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: Past, present and future (a review)," *Digit. Signal Process.*, vol. 106, Nov. 2020, Art. no. 102809.

[4] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, Aug. 2016.

[5] H. Zhi and S. Liu, "Face recognition based on genetic algorithm," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 495–502, Jan. 2019.

[6] J. R. Barr, K. W. Bowyer, P. J. Flynn, and S. Biswas, "Face recognition from video: A review," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 26, no. 5, Aug. 2012, Art. no. 1266002, doi: 10.1142/s0218001412660024.

[7] T. Napoléon and A. Alfalou, "Pose invariant face recognition: 3D model from single photo," *Opt. Lasers Eng.*, vol. 89, pp. 150–161, Feb. 2017.

[8] Y. Guo, H. Wang, L. Wang, Y. Lei, L. Liu, and M. Bennamoun, "3D face recognition: Two decades of progress and prospects," *ACM Comput. Surv.*, vol. 56, no. 3, pp. 1–39, Mar. 2024.

[9] R. A. A. Helmi, S. S. B. E. Yusuf, A. Jamal, and M. I. B. Abdullah, "Face recognition automatic class attendance system (FRACAS)," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (ICACIS)*, Jun. 2019, pp. 50–55.

[10] S. Zhou and S. Xiao, "3D face recognition: A survey," *Hum.-Centric Comput. Inf. Sci.*, vol. 8, no. 1, pp. 1–27, 2018.

[11] X. Yin and X. Liu, "Multi-task convolutional neural network for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 964–975, Feb. 2018.

[12] M. K. Hasan, M. S. Ahsan, S. H. S. Newaz, and G. M. Lee, "Human face detection techniques: A comprehensive review and future research directions," *Electronics*, vol. 10, no. 19, p. 2354, Sep. 2021.

[13] M. Li, B. Huang, and G. Tian, "A comprehensive survey on 3D face recognition methods," *Eng. Appl. Artif. Intell.*, vol. 110, Apr. 2022, Art. no. 104669.

[14] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3D assisted face recognition: A survey of 3D imaging, modelling and recognition approaches," in *Proc. CVPR Workshops*, 2005, p. 114.

[15] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Parana, Brazil, Oct. 2018, pp. 471–478, doi: 10.1109/SIBGRAPI.2018.00067.

[16] Y. Jing, X. Lu, and S. Gao, "3D face recognition: A survey," 2021, *arXiv:2108.11082*.

[17] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, Aug. 2016.

[18] H. Yang and X. A. Wang, "Cascade classifier for face detection," *J. Algorithms Comput. Technol.*, vol. 10, no. 3, pp. 187–197, Sep. 2016.

[19] A. Bansal, C. Castillo, R. Ranjan, and R. Chellappa, "The do's and don'ts for CNN-based face verification," 2017, *arXiv:1705.07426*.

[20] W. Chen, K. Wang, H. Jiang, and M. Li, "Skin color modeling for face detection and segmentation: A review and a new approach," *Multimedia Tools Appl.*, vol. 75, no. 2, pp. 839–862, Jan. 2016, doi: 10.1007/s11042-014-2328-0.

[21] P. Payal and M. M. Goyani, "A comprehensive study on face recognition: Methods and challenges," *Imag. Sci. J.*, vol. 68, no. 2, pp. 114–127, Feb. 2020.

[22] F. Yang, W. Yang, R. Gao, and Q. Liao, "Discriminative multidimensional scaling for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 388–392, Mar. 2018.

[23] M. M. Goyani and N. M. Patel, "Robust facial expression recognition using local Haar mean binary pattern," *J. Inf. Sci. Eng.*, vol. 34, no. 5, pp. 1237–1249, 2018.

[24] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, Sep. 2013.

[25] S. S. Farfade, M. J. Saberian, and L. J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proc. ACM Int. Conf. Multimedia Retrievals*, 2015, pp. 643–650.

[26] C. Peng, X. Gao, N. Wang, and J. Li, "Graphical representation for heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 301–312, Feb. 2017.

[27] J. F. Pereira, R. M. Barreto, G. D. C. Cavalcanti, and I. R. Tsang, "A robust feature extraction algorithm based on class-modular image principal component analysis for face verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1469–1472.

[28] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[29] Y. Kortli, M. Jridi, A. A. Falou, and M. Atri, "A novel face detection approach using local binary pattern histogram and support vector machine," in *Proc. Int. Conf. Adv. Syst. Electric Technol.*, Hammamet, Tunisia, Mar. 2018, pp. 28–33.

[30] Z. Xie, P. Jiang, and S. Zhang, "Fusion of LBP and HOG using multiple kernel learning for infrared face recognition," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 81–84.

[31] L. F. Zhou, Y. W. Du, W. S. Li, J. X. Mi, and X. Luan, "Pose-robust face recognition with Huffman-LBP enhanced by divide-and-rule strategy," *Pattern Recognit.*, vol. 78, pp. 43–55, Jun. 2018.

[32] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," *Exp. Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116743.

[33] M. Karpagam, R. B. Jeyavathana, S. K. Chinnappan, K. V. Kanimozhi, and M. Sambath, "A novel face recognition model for fighting against human trafficking in surveillance videos and rescuing victims," *Soft Comput.*, vol. 27, no. 18, pp. 13165–13180, Sep. 2023.

[34] M. Seshaiah, S. Math, C. Nandini, and M. Rafi, "A googlenet assisted CNN architecture combined with feature attention blocks and Gaussian distribution for video face recognition and verification," *Int. J. Electr. Eng. Technol. (IJEET)*, vol. 12, no. 1, pp. 30–42, 2021.

[35] Z. G. Sheikh and V. M. Thakare, "Wavelet based feature extraction technique for face recognition and retrieval: A review," *Int. Org. Sci. Res. J. Comput. Eng.*, pp. 49–54, 2016. [Online]. Available: www.iosrjournals.org

[36] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 84–91.

[37] A. Ghorbel, I. Tajouri, W. Aydi, and N. Masmoudi, "A comparative study of GOM, uLBP, VLC and fractional eigenfaces for face recognition," in *Proc. Int. Image Process., Appl. Syst. (IPAS)*, Nov. 2016, pp. 1–5.

[38] Z. Li, D. Gong, X. Li, and D. Tao, "Aging face recognition: A hierarchical learning model based on local patterns selection," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2146–2154, May 2016.

[39] K. Vijay and K. Selvakumar, "Brain FMRI clustering using interaction K-means algorithm with PCA," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2015, pp. 909–913.

[40] N. Trendafilov and M. Gallo, "Principal component analysis (PCA)," in *Multivariate Data Analysis on Matrix Manifolds* (Springer Series in the Data Sciences). Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-76974-1_4.

[41] M. H. Mousavi, K. Faez, and A. Asghari, "Three dimensional face recognition using SVM classifier," in *Proc. 7th IEEE/ACIS Int. Conf. Comput. Inf. Sci.*, May 2008, pp. 208–213.

[42] Y. Chu, T. Ahmad, G. Bebis, and L. Zhao, "Low-resolution face recognition with single sample per person," *Signal Process.*, vol. 141, pp. 144–157, Dec. 2017.

[43] E.-J. Cheng, K.-P. Chou, S. Rajora, B.-H. Jin, M. Tanveer, C.-T. Lin, K.-Y. Young, W.-C. Lin, and M. Prasad, "Deep sparse representation classifier for facial recognition and detection system," *Pattern Recognit. Lett.*, vol. 125, pp. 71–77, Jul. 2019.

[44] M. Annalakshmi, S. M. M. Roomi, and A. S. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Comput.*, vol. 22, no. 1, pp. 11–20, Jan. 2019.

[45] J. H. Shah, M. Sharif, M. Raza, and A. Azeem, "A survey: Linear and nonlinear PCA based face recognition techniques," *Int. Arab J. Inf. Technol.*, vol. 10, pp. 536–545, Nov. 2013.

[46] S. R. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2100–2109, Dec. 2014.

[47] Z. Sufyanu, F. S. Mohamad, A. A. Yusuf, and M. B. Mamat, "Enhanced face recognition using discrete cosine transform," *Eng. Lett.*, vol. 24, pp. 52–61, Jan. 2016.

[48] A. Vinay, V. S. Shekhar, K. N. B. Murthy, and S. Natarajan, "Performance study of LDA and KFA for Gabor based face recognition system," *Proc. Comput. Sci.*, vol. 57, pp. 960–969, Jan. 2015.

[49] Y. Wang, P. Fei, X. Fan, and H. Li, "Face recognition using nonlinear locality preserving with deep networks," in *Proc. 7th Int. Conf. Internet Multimedia Comput. Service*, Hunan, China, Aug. 2015, p. 66.

[50] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[51] H. Zhao, P. Chi Yuen, and J. T. Kwok, "A novel incremental principal component analysis and its application for face recognition," *IEEE Trans. Syst., Man Cybern., B*, vol. 36, no. 4, pp. 873–886, Aug. 2006.

[52] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, pp. 1–6.

[53] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[54] H. Dibeklioglu, F. Alnajar, A. Ali Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, Jun. 2015.

[55] A. Chehrehgosha and M. Emadi, "Face detection using fusion of LBP and AdaBoost," *J. Soft Comput. Appl.*, vol. 2016, no. 1, pp. 1–10, 2016.

[56] C. A. Kumar and D. Maitreyee, "Face detection and recognition using local binary patterns," *Int. J. Adv. Res. Elect., Electron. Instrum. Eng.*, vol. 5, no. 10, pp. 7923–7929, Oct. 2016.

[57] G. Ahmed and B. Mohamed, "Improving the recognition of faces using PCA and SVM optimized by DWT," *Int. J. Comput. Appl.*, vol. 107, no. 17, pp. 7–11, Dec. 2014.

[58] S. Mousavi, M. Charmi, and H. Hassanpoor, "Recognition of identical twins based on the most distinctive region of the face: Human criteria and machine processing approaches," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15765–15802, Apr. 2021.

[59] S. Mousavi, M. Charmi, and H. Hassanpoor, "A distinctive landmark-based face recognition system for identical twins by extracting novel weighted features," *Comput. Electr. Eng.*, vol. 94, Sep. 2021, Art. no. 107326, doi: 10.1016/j.compeleceng.2021.107326.

[60] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq, "3D face recognition using covariance based descriptors," *Pattern Recognit. Lett.*, vol. 78, pp. 1–7, Jul. 2016.

[61] H. Boukamcha, M. Elhallek, M. Atri, and F. Smach, "3D face landmark auto detection," in *Proc. World Symp. Comput. Netw. Inf. Secur. (WSCNIS)*, Hammamet, Tunisia, Sep. 2015, pp. 1–6.

[62] M. K. Hasan, M. S. Ahsan, S. S. Newaz, and G. M. Lee, "Human face detection techniques: A comprehensive review and future research directions," *Electronics*, vol. 10, no. 19, p. 2354, 2021.

[63] N. A. Talemi, H. Kashiani, and N. M. Nasrabadi, "CATFace: Cross-attribute-guided transformer with self-attention distillation for low-quality face recognition," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 6, no. 1, pp. 132–146, Jan. 2024.

[64] A. A. Fathima, S. Ajitha, V. Vaidehi, M. Hemalatha, R. Karthigaiveni, and R. Kumar, "Hybrid approach for face recognition combining Gabor wavelet and linear discriminant analysis," in *Proc. IEEE Int. Conf. Comput. Graph., Vis. Inf. Secur. (CGVIS)*, Bhubaneswar, India, Nov. 2015, pp. 220–225.

[65] C.-H. Hu, X.-B. Lu, P. Liu, X.-Y. Jing, and D. Yue, "Single sample face recognition under varying illumination via QRCP decomposition," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2624–2638, May 2019.

[66] Y. Yan, H. Wang, and D. Suter, "Multi-subregion based correlation filter bank for robust face recognition," *Pattern Recognit.*, vol. 47, pp. 3487–3501, Nov. 2014.

[67] Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Phillips, 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, 2016.

[68] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14220–14229.

[69] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805.

[70] J. Sun, Y. Fu, S. Li, J. He, C. Xu, and L. Tan, "Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors," *J. Sensors*, vol. 2018, pp. 1–10, Sep. 2018.

[71] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[72] F. Fanaee, M. Yazdi, and M. Faghihi, "Face image super-resolution via sparse representation and wavelet transform," *Signal, Image Video Process.*, vol. 13, no. 1, pp. 79–86, Feb. 2019.

[73] U. Sharma, K. N. Faisal, R. R. Sharma, and K. V. Arya, "Facial landmark-based human emotion recognition technique for oriented viewpoints in the presence of facial attributes," *Social Netw. Comput. Sci.*, vol. 4, no. 3, p. 273, Mar. 2023.

[74] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," 2019, *arXiv:1907.06724*.

[75] M. Taskiran, N. Kahraman, and C. E. Erdem, "Hybrid face recognition under adverse conditions using appearance-based and dynamic features of smile expression," *IET Biometrics*, vol. 10, no. 1, pp. 99–115, Jan. 2021.

[76] M. Z. Khan, S. Harous, S. U. Hassan, M. U. G. Khan, R. Iqbal, and S. Mumtaz, "Deep unified model for face recognition based on convolution neural network and edge computing," *IEEE Access*, vol. 7, pp. 72622–72633, 2019.

[77] S. Li, L. Huang, W. Zhang, and B. Tang, "Hybrid feature fusion for person recognition in photo albums," in *Proc. MIPPR Pattern Recognit. Comput. Vis.*, Feb. 2020, Art. no. 114300.

[78] S. A. A. Shah, W. Deng, M. A. Cheema, and A. Bais, "CommuNety: Deep learning-based face recognition system for the prediction of cohesive communities," *Multimedia Tools Appl.*, vol. 82, no. 7, pp. 10641–10659, Mar. 2023, doi: 10.1007/s11042-022-13741-y.

[79] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2781–2794, Nov. 2020, doi: 10.1109/TPAMI.2019.2914680.

[80] S. Z. Gilani, A. Mian, and P. Eastwood, "Deep, dense and accurate 3D face correspondence for generating population specific deformable models," *Pattern Recognit.*, vol. 69, pp. 238–250, Sep. 2017.

[81] M. He, J. Zhang, S. Shan, and X. Chen, "Enhancing face recognition with detachable self-supervised bypass networks," *IEEE Trans. Image Process.*, vol. 33, pp. 1588–1599, 2024.

[82] M. Rostami, A. Farajollahi, and H. Parvin, "Deep learning-based face detection and recognition on drones," *J. Ambient Intell. Humanized Comput.*, vol. 15, no. 1, pp. 373–387, Jan. 2024.

[83] M. Sohail, I. A. Shoukat, A. U. Khan, H. Fatima, M. R. Jafri, M. A. Yaqub, and A. Liotta, "Deep learning based multi pose human face matching system," *IEEE Access*, vol. 12, pp. 26046–26061, 2024.

[84] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 18729–18738.

[85] G. Sunitha, K. Geetha, S. Neelakandan, A. K. S. Pundir, S. Hemalatha, and V. Kumar, "Intelligent deep learning based ethnicity recognition and classification using facial images," *Image Vis. Comput.*, vol. 121, May 2022, Art. no. 104404, doi: 10.1016/j.imavis.2022.104404.

[86] M. Villegas and R. Paredes, "On improving robustness of LDA and SRDA by using tangent vectors," *Pattern Recognit. Lett.*, vol. 34, no. 9, pp. 1094–1100, Jul. 2013.
[87] A.-K. Seghouane, N. Shokouhi, and I. Koch, "Sparse principal component analysis with preserved sparsity pattern," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3274–3285, Jul. 2019.
[88] R. Martin-Clemente and V. Zarzoso, "On the link between L1-PCA and ICA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 515–528, Mar. 2017.
[89] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.
[90] C.-N. Li, Y.-H. Shao, W. Yin, and M.-Z. Liu, "Robust and sparse linear discriminant analysis via an alternating direction method of multipliers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 915–926, Mar. 2020.
[91] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[92] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
[93] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
[94] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 1–8.
[95] D. Cheng, X. Zhang, and X. Xu, "Reweighted robust and discriminative latent subspace projection for face recognition," *Inf. Sci.*, vol. 657, Feb. 2024, Art. no. 119947.
[96] Z. Yang, X. Wu, P. Huang, F. Zhang, M. Wan, and Z. Lai, "Orthogonal autoencoder regression for image classification," *Inf. Sci.*, vol. 618, pp. 400–416, Dec. 2022.
[97] N. D. Kalka, B. Maze, J. A. Duncan, K. O'Connor, S. Elliott, K. Hebert, J. Bryan, and A. K. Jain, "IJB-S: IARPA Janus surveillance video benchmark," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–9.
[98] J. Zheng and R. Chellappa, "Video-based face recognition," in *Computer Vision*, K. Ikeuchi, Ed., Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-63416-2_816.
[99] T. E. de Campos, R. S. Feris, and R. M. Cesar Jr., "A framework for face recognition from video sequences using GWN and eigenfeature selection," in *Proc. Workshop Artif. Intell. Comput. Vis.*, 2000, pp. 1–10.
[100] V. Kruger, A. Happe, and G. Sommer, "Affine real-time face tracking using wavelet network," in *Proc. Int. Workshop Recognit., Anal., Tracking Faces Gestures Real-Time Syst.*, Corfu, Greece, 1999, pp. 141–148.
[101] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. BMVC*, 2015, pp. 1–12.
[102] R. Wang, S. Shan, X. Chen, Q. Dai, and W. Gao, "Manifold–manifold distance and its application to face recognition with image sets," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4466–4479, Oct. 2012.
[103] J. Zheng, R. Ranjan, C.-H. Chen, J.-C. Chen, C. D. Castillo, and R. Chellappa, "An automatic system for unconstrained video-based face recognition," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 3, pp. 194–209, Jul. 2020.
[104] V. Kruger and S. Zhou, "Exemplar-based face recognition from video," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Florence, Italy, Oct. 2012, pp. 766–779.
[105] S. Gong, Y. Shi, N. D. Kalka, and A. K. Jain, "Video face recognition: Component-wise feature aggregation network (C-FAN)," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.
[106] M. Paleari, C. Velardo, B. Huet, and J.-L. Dugelay, "Face dynamics for biometric people recognition," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Oct. 2009, pp. 1–5.
[107] S. Gong, Y. Shi, and A. K. Jain, "Recurrent embedding aggregation network for video face recognition," 2019, *arXiv:1904.12019*.
[108] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, "Pain expression as a biometric: Why patients' self-reported pain doesn't match with the objectively measured pain?" in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Feb. 2017, pp. 1–8.
[109] M. Sharif, F. Naz, M. Yasmin, M. A. Shahid, and A. Rehman, "Face recognition: A survey," *J. Eng. Sci. Technol. Rev.*, vol. 10, no. 2, pp. 166–177, 2017.
[110] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.
[111] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, Jun. 2011, pp. 529–534, doi: 10.1109/CVPR.2011.5995566.
[112] A. Georghiades, P. Belhumeur, and D. Kriegman. (1997). *Yale Face Database*. [Online]. Available: http://cvc.yale.edu/projects/yalefaces/yalefa
[113] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*.
[114] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Proc. ECCV*, 2016, pp. 87–102.
[115] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021.
[116] A. Moreno, "GavabDB: A 3D face database," in *Proc. 2nd COST Workshop Biometrics Internet*, 2004, pp. 75–80.
[117] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother, "IARPA Janus benchmark-B face dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 592–600.
[118] C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps," *J. Real-Time Image Process.*, vol. 12, no. 1, pp. 155–163, Jun. 2016.
[119] X. Deng, F. Da, and H. Shao, "Efficient 3D face recognition using local covariance descriptor and Riemannian kernel sparse coding," *Comput. Electr. Eng.*, vol. 62, pp. 81–91, Aug. 2017.
[120] A. Abbad, K. Abbad, and H. Tairi, "3D face recognition: Multi-scale strategy based on geometric and local descriptors," *Comput. Electr. Eng.*, vol. 70, pp. 525–537, Aug. 2018.
[121] L. Shi, X. Wang, and Y. Shen, "Research on 3D face recognition method based on LBP and SVM," *Optik*, vol. 220, Oct. 2020, Art. no. 165157.
[122] Y. Liang, Y. Zhang, and X.-X. Zeng, "Pose-invariant 3D face recognition using half face," *Signal Process., Image Commun.*, vol. 57, pp. 84–90, Sep. 2017.
[123] K. Papadopoulos, A. Kacem, A. Shabayek, and D. Aouada, "Face-GCN: A graph convolutional network for 3D dynamic face identification/recognition," 2021, *arXiv:2104.09145*.
[124] A. Verma, A. Goyal, N. Kumar, and H. Tekchandani, "Face recognition: A review and analysis," in *Proc. ICCIDM*, 2022, pp. 195–210.
[125] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: A complete review," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 4825–4880, Jan. 2021.

H. L. GURURAJ (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. He is currently an Associate Professor with the Department of Information Technology, Manipal Institute of Technology, Bengaluru, India. He is also the Founder of the Wireless Internetworking Group (WiNG). He has 13 years of teaching experience at both UG and PG levels. He has guided 30 UG students and 20 PG students. He has published more than 150 research articles, including 20 SCI publications. He has edited 20 books at Springer, IET, IGI Global, and Taylor and Francis. He has presented 55 papers at various international conferences. His research interests include blockchain technology, cyber security, wireless sensor networks, ad-hoc networks, the IoT, data mining, cloud computing, and machine learning. He is a Senior Member of ACM and an ACM Distinguished Speaker. He is a Lifetime Member of ISTE and CSI. He received the Young Scientist International Travel Grant Award from the Government of India.

B. C. SOUNDARYA received the M.Tech. degree in CSE from VVCE, Mysuru. She is currently a Researcher with Manipal Institute of Technology, Bengaluru. She has published various research articles in national and international journals. Her research interests include biometrics, cyber security, machine learning, and deep learning. She is a member of ACM and an ACM Student Advisor. She has received best paper awards at various conferences.

S. PRIYA received the B.E. and M.E. degrees (Hons.) in computer science and engineering from Anna University, Chennai, India, in 2009 and 2013, respectively, and the Doctor of Philosophy degree from the National Institute of Technology, Tiruchirappalli, in 2022. She has six years of teaching experience. She is currently an Assistant Professor with the Department of Information Technology, Manipal Institute of Technology, Bengaluru, Karnataka. She published her research papers in refereed international journals with high impact factors and international conferences. Her research interests include mobile security, network security, social network attack detection, evolutionary algorithms, and neural networks.

J. SHREYAS received the Ph.D. degree in computer science and engineering from Bengaluru University, Belagavi, India. He is currently an Assistant Professor with the Department of Information Technology, Manipal Institute of Technology, Bengaluru, India. He has four years of teaching experience at both UG and PG levels. He has published more research articles, including SCI publications. He has edited books at Springer, IET, IGI Global, and Taylor and Francis. His research interests include blockchain technology, cyber security, wireless sensor networks, ad-hoc networks, the IoT, data mining, cloud computing, and machine learning. He is a member of ACM.

FRANCESCO FLAMMINI (Senior Member, IEEE) was born in Formia, Italy, in 1978. He received the master's degree (cum laude) in computer engineering and the Ph.D. degree in computer and automation engineering from the University of Naples Federico II, Italy, in 2003 and 2006, respectively. From 2003 to 2007, he was a Software V&V Engineer with Ansaldo STS (now Hitachi Rail), where he was a Senior Innovation Engineer, from 2007 to 2016. From 2016 to 2017, he was an Information Security Compliance Manager with Italian State Mint and Polygraphic Institute. Since 2018, he has been an Associate Professor with Linnaeus University, Sweden, where he has chaired the cyber-physical systems (CPS) environment. Since 2020, he has been a Professor of computer science with Mälardalen University, Sweden. He had leadership roles in more than ten research projects. He has edited or authored more than ten books and 130 publications. His research interests include resilient cyber-physical systems and trustworthy autonomy. He is an ACM Distinguished Speaker and the Chair of the IEEE SMC Technical Committee on Homeland Security.