Past, present, and future of face recognition: a review

Insaf Adjabi, Abdeldjalil Ouahabi, Amir Benzaoui, Abdelmalik Taleb-Ahmed

electronics

MDPI

Review

Past, Present, and Future of Face Recognition: A Review

Insaf Adjabi [1], Abdeldjalil Ouahabi [1,2,*], Amir Benzaoui [3] and Abdelmalik Taleb-Ahmed [4]

[1] Department of Computer Sciences, LIMPAF, University of Bouira, Bouira 10000, Algeria; i.adjabi@univ-bouira.dz
[2] Polytech Tours, Imaging and Brain, INSERM U930, University of Tours, 37200 Tours, France
[3] Department of Electrical Engineering, University of Bouira, Bouira 10000, Algeria; a.benzaoui@univ-bouira.dz
[4] Laboratory of IEMN DOAE. UMR CNRS 8520, University of Valenciennes, 59313 Valenciennes, France; Abdelmalik.Taleb-Ahmed@uphf.fr
* Correspondence: ouahabi@univ-tours.fr

Received: 16 June 2020; Accepted: 18 July 2020; Published: 23 July 2020

Abstract: Face recognition is one of the most active research fields of computer vision and pattern recognition, with many practical and commercial applications including identification, access control, forensics, and human-computer interactions. However, identifying a face in a crowd raises serious questions about individual freedoms and poses ethical issues. Significant methods, algorithms, approaches, and databases have been proposed over recent years to study constrained and unconstrained face recognition. 2D approaches reached some degree of maturity and reported very high rates of recognition. This performance is achieved in controlled environments where the acquisition parameters are controlled, such as lighting, angle of view, and distance between the camera–subject. However, if the ambient conditions (e.g., lighting) or the facial appearance (e.g., pose or facial expression) change, this performance will degrade dramatically. 3D approaches were proposed as an alternative solution to the problems mentioned above. The advantage of 3D data lies in its invariance to pose and lighting conditions, which has enhanced recognition systems efficiency. 3D data, however, is somewhat sensitive to changes in facial expressions. This review presents the history of face recognition technology, the current state-of-the-art methodologies, and future directions. We specifically concentrate on the most recent databases, 2D and 3D face recognition methods. Besides, we pay particular attention to deep learning approach as it presents the actuality in this field. Open issues are examined and potential directions for research in facial recognition are proposed in order to provide the reader with a point of reference for topics that deserve consideration.

Keywords: face recognition; face analysis; face database; deep learning

1. Introduction

Face recognition has gained tremendous attention over the last three decades since it is considered a simplified image analysis and pattern recognition application. There are at least two reasons for understanding this trend: (1) the large variety of commercial and legal requests, besides (2) the availability of the relevant technologies (e.g., smartphones, digital cameras, GPU, ...). Although the existing machine learning/recognition systems have achieved some degree of maturity, their performance is limited to the conditions imposed in real-world applications [1]. For example, identifying facial images obtained in an unconstrained environment (e.g., changes in lighting, posture, or facial expression, in addition to partial occlusion, disguises, or camera movement) still poses several challenges ahead. In other words, the existing technologies are still far removed from the human visual system capabilities.

In our daily lives, the face is perhaps the most common and familiar biometric feature. With the invention of photography, government departments and private entities have kept facial photographs (from personal identity documents, passports, or membership cards). These collections have been used in forensic investigations, as referential databases, to match and compare a respondent's facial images (e.g., perpetrator, witness, or victim). Besides, the broad use of digital cameras and smartphones made facial images easy to produce every day; these images can be easily distributed and exchanged by rapidly established social networks such as Facebook and Twitter.

Face recognition has a long history; it stirs neurologists, psychologists, and computer scientists [2]. The human face is not an ideal modality compared to other biometric traits; it is typically less precise than other biometric modalities such as iris or fingerprint, and can potentially be influenced by cosmetics, disguises, and lighting [3]. However, the face has the advantages that make it one of the most favored biometric characteristics for identity recognition, we can note:

- Natural character: The face is a very realistic biometric feature used by humans in the individual's recognition, making it possibly the most related biometric feature for authentication and identification purposes [4]. For example, in access control, it is simple for administrators to monitor and evaluate approved persons after authentication, using their facial characteristics. The support of ordinary employers (e.g., administrators) may boost the efficiency and applicability of recognition systems. On the other hand, identifying fingerprints or iris requires an expert with professional competencies to provide accurate confirmation.
- Nonintrusive: In contrast to fingerprint or iris images, facial images can quickly be obtained without physical contact; people feel more relaxed when using the face as a biometric identifier. Besides, a face recognition device can collect data in a friendly manner that people commonly accept [5].
- Less cooperation: Face recognition requires less assistance from the user compared with iris or fingerprint. For some limited applications such as surveillance, a face recognition device may recognize an individual without active subject involvement [5].

First attempts at identifying a facial subject by comparing a part of a facial photograph were reported at a British court in 1871 [6]. Face recognition is one of the most significant law enforcement techniques in cases where video material or pictures on a crime scene are available. Legal specialists do a manual facial image test to match that of a suspect. Automated facial recognition technologies have increased the efficiency of judicial employees and streamlined the comparison process [7].

Today facial recognition, associated with artificial intelligence techniques, enables a person to be identified from his face or verified as what he claims to be. Facial recognition can analyze facial features and other biometric details, such as the eyes, and compare them with photographs or videos. With accusations of widespread surveillance, this controversial technology raises many concerns among its opponents, who fear breaches of data privacy and individual liberties. Face recognition for its defenders enables accurate, fast, and secure authentication to protect against all fraud forms. According to a report by the analytical company Mordor-Intelligence [8], the face recognition market was estimated at 4.4 billion dollars worldwide in 2019 and would surpass 10.9 billion in 2025. This technology has already become popular in some countries, such as China.

Because of artificial intelligence technologies, significant advances in face recognition have occurred. In early times, research interests were mainly focused on face recognition under controlled conditions where simple classical approaches provided excellent performance. Today, the focus of research is on unconstrained conditions in which deep learning technology [9] has gained more popularity as it offers strong robustness against the numerous variations that can alter the recognition process.

In addition, many academics struggle to find robust and reliable data sets for testing and to evaluate their proposed method: finding an appropriate data set is an important challenge especially in 3D facial recognition and facial expression recognition. To check the effectiveness of these methods,

accurate datasets are required that (i) contain a large number of persons and photographs, (ii) follow real-world requirements, and (iii) are open to the public.

Our contributions in this review are:

- We provide an updated review of automated face recognition systems: the history, present, and future challenges.
- We present 23 well-known face recognition datasets in addition to their assessment protocols.
- We have reviewed and summarized nearly 180 scientific publications on facial recognition and its material problems of data acquisition and pre-processing from 1990 to 2020. These publications have been classified according to various approaches: holistic, geometric, local texture, and deep learning for 2D and 3D facial recognition. We pay particular attention to the methods based deep-learning, which are currently considered state-of-the-art in 2D face recognition.
- We analyze and compare several in-depth learning methods according to the architecture implemented and their performance assessment metrics.
- We study the performance of deep learning methods under the most commonly used data set: (i) Labeled Face in the Wild (LFW) data set [10] for 2D face recognition, (ii) Bosphorus and BU-3DFE for 3D face recognition.
- We discuss some new directions and future challenges for facial recognition technology by paying particular attention to the aspect of 3D recognition.

2. Face Recognition History

This section reviews the most significant historical stages that have contributed to the advancement of face recognition technology (outlined in Figure 1):

- 1964: The American researchers Bledsoe et al. [11] studied facial recognition computer programming. They imagine a semi-automatic method, where operators are asked to enter twenty computer measures, such as the size of the mouth or the eyes.
- 1977: The system was improved by adding 21 additional markers (e.g., lip width, hair color).
- 1988: Artificial intelligence was introduced to develop previously used theoretical tools, which showed many weaknesses. Mathematics ("linear algebra") was used to interpret images differently and find a way to simplify and manipulate them independent of human markers.
- 1991: Alex Pentland and Matthew Turk of the Massachusetts Institute of Technology (MIT) presented the first successful example of facial recognition technology, Eigenfaces [12], which uses the statistical Principal component analysis (PCA) method.
- 1998: To encourage industry and the academy to move forward on this topic, the Defense Advanced Research Projects Agency (DARPA) developed the Face recognition technology (FERET) [13] program, which provided to the world a sizable, challenging database composed of 2400 images for 850 persons.
- 2005: The Face Recognition Grand Challenge (FRGC) [14] competition was launched to encourage and develop face recognition technology designed to support existent facial recognition initiatives.
- 2011: Everything accelerates due to deep learning, a machine learning method based on artificial neural networks [9]. The computer selects the points to be compared: it learns better when it supplies more images.
- 2014: Facebook knows how to recognize faces due to its internal algorithm, Deepface [15]. The social network claims that its method approaches the performance of the human eye near to 97%.

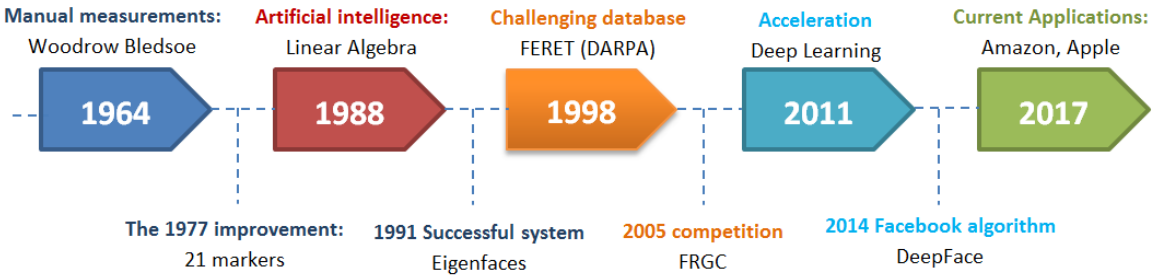


Figure 1. Primary stages in the history of face recognition.

Today, facial recognition technology advancement has encouraged multiple investments in commercial, industrial, legal, and governmental applications. For example:

- In its new updates, Apple introduced a facial recognition application where its implementation has extended to retail and banking.
- Mastercard developed the Selfie Pay, a facial recognition framework for online transactions.
- From 2019, people in China who want to buy a new phone will now consent to have their faces checked by the operator.
- Chinese police used a smart monitoring system based on live facial recognition; using this system, they arrested, in 2018, a suspect of "economic crime" at a concert where his face, listed in a national database, was identified in a crowd of 50,000 persons.

3. Face Recognition Systems

3.1. Main Steps in Face Recognition Systems

In engineering, the issue of automated face recognition includes three key steps [16] (as presented in Figure 2): (1) approximate face detection and normalization, (2) extraction of features and accurate face normalization, and (3) classification (verification or identification).

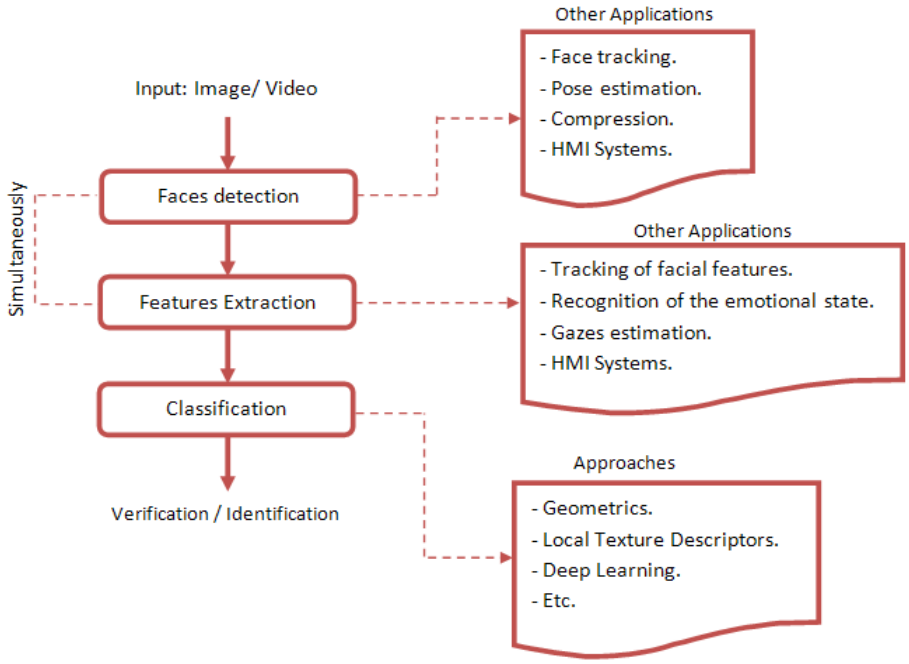


Figure 2. The standard design of an automated face-recognition system.

Face detection is the first step in the automated face recognition system. It usually determines whether or not an image includes a face(s). If it does, its function is to trace one or several face locations in the picture [17].

Feature extraction step consists of extracting from the detected face a feature vector named the signature, which must be enough to represent a face. The individuality of the face and the property of distinguishing between two separate persons must be checked. It should be noted that the face detection stage can accomplish this process.

Classification involves verification and identification. Verification requires matching one face to another to authorize access to a requested identity. However, identification compares a face to several other faces that are given with several possibilities to find the face's identity.

Sometimes, some steps are not separated. For example, the facial features (eyes, mouth, and nose) used for feature extraction are frequently used during face detection. Detection and extraction of features can be performed simultaneously, as shown in Figure 2.

Depending on the application environment's complexity, some external factors can cause highly intra-face identity distributions (or lowly inter-face identity distributions) and degrade the accuracy of recognition. Among these factors, we can cite the database size, low or high lighting, presence of noise or blur, disguises, partial occlusion, and certain secondary factors that are often common, unavoidable, and very challenging [18]. In a noisy environment, image pre-processing may prove necessary [19–21].

Although automated face recognition systems must perform the three steps mentioned above, each step is considered a critical research issue, not only because the techniques used for each step need to be improved and because they are essential in several applications, as shown in Figure 2. For example, face detection is necessary to activate facial monitoring, and the extraction of facial features is crucial to identify the person's emotional state, which is, in turn, essential in human–machine interaction systems (HMI). The isolation of each step facilitates the evaluation and state-of-the-art evolution.

This paper mainly focuses on feature extraction (and possibly feature selection) and classification. The face acquisition and detection step is a critical problem analyzed in the case of 3D facial recognition.

3.2. Assessment Protocols in Face Recognition

As stated in the previous sub-section, an automated face recognition system can operate either in the mode of verification or identification, depending on each application (as seen in Figure 3).

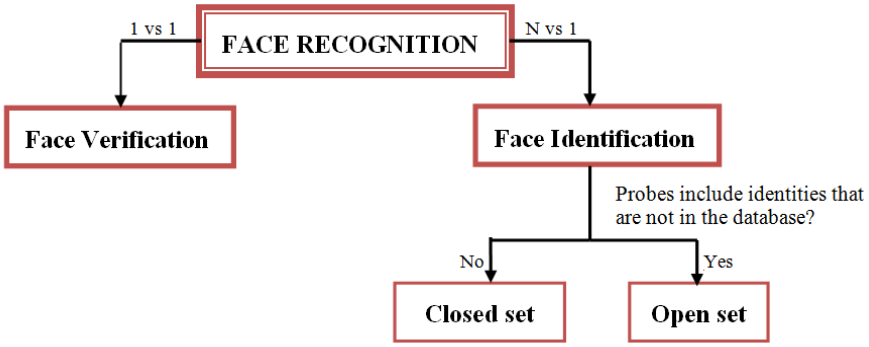


Figure 3. Categorization of various assessment protocols in face recognition.

In verification mode [22], the system evaluates a person's identity by comparing his/her registered model(s) in the database with the captured face. A one-to-one comparison is performed by the system to decide whether the proclaimed identity is true or false. Habitually, verification is used for positive recognition to avoid different individuals using the same identity. Face verification systems are classically assessed by the receiver operating characteristic (*ROC*) and the estimated mean accuracy (*ACC*).

Two types of errors are assessed for *ROC* analysis: true accept rate (*TAR*) and false accept rate (*FAR*). The *TAR* is defined as the fraction of valid comparisons exceeding the similarity score (threshold) correctly:

$$TAR = \frac{TP}{(TP + FN)} \tag{1}$$

TP: true positive.

FN: false negative.

Moreover, *FAR* is defined as the fraction of the impostor comparisons exceeding incorrectly the same threshold:

$$FAR = \frac{FP}{(FP + TN)} \tag{2}$$

FP: false positive.

TN: true negative.

However, *ACC* is a simplified metric, which shows the percentage of correct classifications:

$$ACC = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{3}$$

In identification mode [22], the system identifies an individual by searching for the enrolled model representing the best match between all facial models stored in the database. Therefore, a one-against-all comparison is performed by the system to determine this individual (or failure if that individual does not exist in the database), without providing a prior declaration of identity.

Identification is an essential task for harmful recognition applications; the purpose of this type of recognition is to prevent multiple identities by one single individual. For two different scenarios, two test protocols may be used, which are: open-set and closed-set (as shown in Figure 3).

For the former, the training set cannot include test identities. Different metrics are established in the open-set face identification scenario to measure the model's accuracy such as the false negative identification rate (FNIR) and the false positive identification rate (FPIR). FNIR measures the ratio of cases wrongly classified as false, although they are true cases, while FPIR measures the ratio of cases wrongly classified as true despite being false.

Whereas the latter retrieves images from the same identities for training and testing. Rank-N is a fundamental performance metric used in closed-set face identification to measure the model's accuracy, where the valid user identifier is returned within the N-Top matches. The primary measuring performance is recorded using correct identification rates on a cumulative match characteristics (CMC) curve.

4. Available Datasets and Protocols for 2D Face Recognition

To evaluate and compare the verification or identification performance of a pattern recognition system, in general, and a biometric recognition system in particular, image benchmark datasets of adequate subject size must be accessible to the public. In this section, we would like summarize appropriate and recent datasets for testing the performance of face verification and identification systems, which can also be freely downloaded or certified with an acceptable effort. We concentrate primarily on the datasets that are only appropriate for testing approaches to 2D face recognition. Figure 4 summarizes the datasets listed in the chronological order of their appearance.

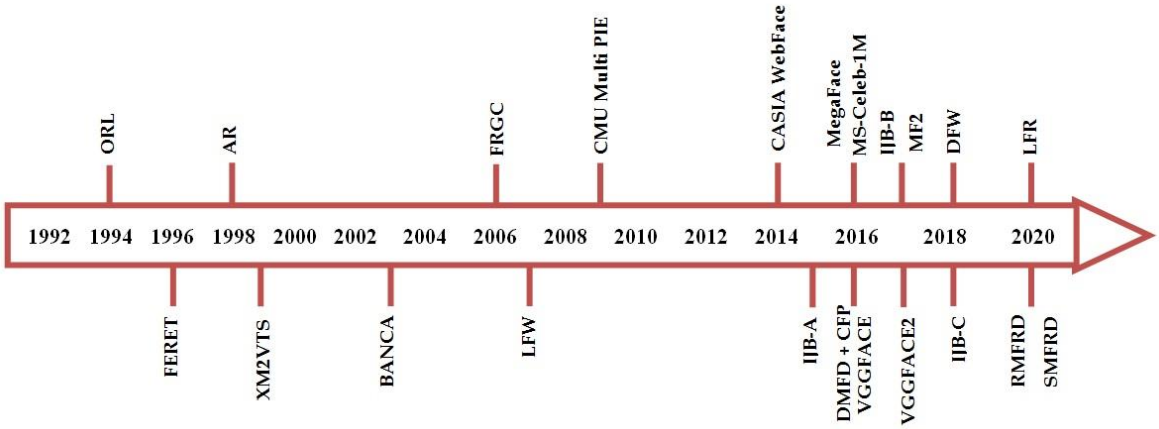


Figure 4. The developments of 2D face recognition datasets through time.

4.1. ORL Dataset

Between 1992 and 1994, the ORL (Olivetti Research Laboratory) dataset [23] was developed at the Cambridge University Computer Laboratory. It was employed in the framework of a face recognition project carried out in cooperation with the Speech, Vision, and Robotics Group of the Cambridge University Engineering Department. It contains 400 frontal facial images for 40 persons with different facial expressions (open/closed eyes, smiling/not smiling), conditions of illumination, hairstyles with or without the beard, mustaches, and glasses. All the samples were acquired against a dark uniform background with the subjects in a vertical frontal pose (with a little side movement). Each sample is 92×112 pixels with 256 grayscale images, with tilting tolerance and up to 20° rotation, as shown in Figure 5.

Figure 5. Some samples from the ORL (Olivetti Research Laboratory) database.

4.2. FERET Dataset

The FERET (Facial recognition technology) dataset [13] was created in the Department of Defense Counterdrug Technology Development Program Office between 1993 and 1996. The goal was to develop face recognition capabilities through machines that could be used to assist authentication, security, and forensic applications. The facial images were acquired in 15 sessions in a semi-controlled environment. It covers 1564 sets of facial images for 14,126 images that comprise 1199 persons and 365 duplicate sets of facial images (Figure 6). A duplicate set is another set of facial images of an individual already existing in the database and was typically collected on another day.

Figure 6. Some samples from the face recognition technology (FERET) database.

The most commonly used evaluation protocol requires algorithms matching a set of 3323 samples against 3816 samples, performing about 12.6 million matches. Besides, it allows the determination of recognition scores for various galleries and probe sets. The following cases can be studied: (1) The gallery and probe samples of an individual were acquired on the same day, under the same illumination situations; (2) the gallery and probe samples of a person were acquired on several days; (3) the gallery and probe samples of a person were acquired over a year apart; and (4) the gallery and probe samples of a person were acquired on the same day, but with diverse illumination conditions.

The protocol offers two types of evaluation: the first provides the samples with the eyes center coordinates, and the second provides the samples only without any indication.

4.3. AR Dataset

Martinez and Benavente produced the database AR (Alex and Robert) [24] in 1998 at the Computer Vision Center, Barcelona (Spain). The database includes more than 3000 colored facial images of 116 subjects (53 women and 63 men). This database's images were taken under different illumination conditions, with frontal views, facial expressions, and occlusions (by scarf and sun-glasses). All images were acquired under rigorously controlled situations: no wear requirements (glasses, clothes), hair-style, and make-up were required for contributors. Each individual took part in two separate sessions, spaced up by two weeks. So, there are 26 images from each subject. Both sessions acquired the same images. The resulting red, green and blue (RGB) facial images are 768×576 pixels in size. The database provides 13 kinds of images (Figure 7) which are: (1) neutral expression, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both side-lights on, (8) wearing sunglasses, (9) wearing sunglasses and left light on, (10) wearing sunglasses and right light on, (11) wearing a scarf, (12) wearing scarf and left a light on, and (13) wearing scarf and right light on.

Figure 7. Examples from the AR (Alex and Robert) database.

4.4. XM2VTS Database

The XM2VTS (multi modal verification for teleservices and security applications) database [25] contains facial videos from 295 persons. The images were acquired at a one-month interval spaced over four different sessions, and the subject is composed of a group of 200 training participants, 25 evaluation impostors, and 70 check impostors. Figure 8 shows two examples of one shot from each session for a dataset person. Both configurations of Lausanne protocol LPI and LPII are fixed for XM2VTS to evaluate the verification mode's performance. The database is decomposed into three groups: train, evaluation, and test: The training group assists in model estimation; the evaluation group is employed to adjust system parameters and system performance can be measured on the test group using parameters of evaluation. The difference between the LPI and LPII configurations is the number of facial samples in each group.

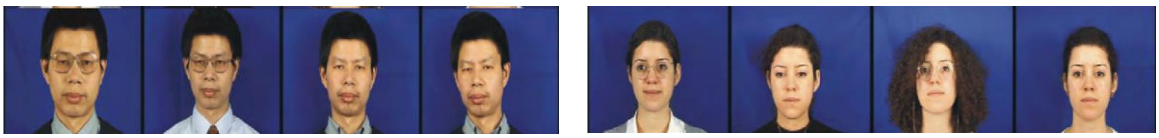


Figure 8. Examples of XM2VTS (multi modal verification for teleservices and security applications) facial images of the same subject under different periods.

4.5. BANCA Dataset

The BANCA dataset [26] is a tremendous, practical, and challenging multi-modal dataset proposed in 2003 for training and testing multi-modal biometric verification systems. It was acquired in four European languages, which offer two modalities: voice and face. Both high and low-quality cameras and microphones were employed for acquisition. The images/voices were acquired in three diverse scenarios (controlled, degraded, and adverse) over 12 different sessions in three months. Totalities of 208 persons were acquired: 104 men and 104 women. Figure 9 shows some facial examples.

(**a**) (**b**) (**c**)

Figure 9. Facial examples from the BANCA database: (**a**) controlled, (**b**) degraded, and (**c**) adverse.

Seven separate experimental configurations were established for the evaluation protocol to determine which material should be employed for training and which one can use for testing. Such configurations include: (1) matched controlled (MC), (2) matched degraded (MD), (3) matched adverse (MA), (4) unmatched degraded (UD), (5) unmatched adverse (UA), (6) pooled test (P), and (7) grand test (G). Note that each person should be trained to employ the face images from the controlled scenario's first recording session.

4.6. FRGC Dataset

From 2004 to 2006, the Face Recognition Grand Challenge (FRGC) [14] was produced at the University of Notre Dame to achieve performance improvements by pursuing algorithm progress for all methods proposed in the literature. The FRGC dataset contains 50,000 images that are decomposed into training and validation parts. A subject session in FRGC consists of four controlled still images, two uncontrolled images, and one 3D image, as shown in Figure 10. The FRGC is distributed with six experimental protocols:

- In experimental protocol 1, two controlled still images of an individual are used as one for a gallery, and the other for a probe.
- In Exp 2, the four controlled images of a person are distributed among the gallery and probe.
- In Exp 4, a single controlled still image presents the gallery, and a single uncontrolled still image presents the probe.
- Exps 3, 5, and 6 are designed for 3D images.

Figure 10. Samples from one person session: controlled stills, uncontrolled stills, and 3D shape.

4.7. LFW Database

The facial image database: Labeled Faces in the Wild (LFW), created by Huang et al. [10] in 2007, desired to study the unconstrained problem of face recognition, such as variation in posture, facial expression, race, background, ethnicity, lighting, gender, age, color saturation, clothing, camera quality, hairstyles, focus, and other parameters (Figure 11). It contains 13,233 facial images collected from the web: 5749 individuals where 1680 have two or more distinct images. Each image is 250 × 250 pixels in size. Two protocols are defined for using the training-data: Image-restricted and unrestricted training. Under image-restricted, the identities of images are ignored in training. The principal difference is that the unrestricted protocol permits to form as many impostors and genuine pairs as possible over the restricted training pairs [10]. Three types of aligned images are proposed: (1) the funneled images [27], (2) LFW-A used an unpublished method for alignment, and (3) deep funneled images [28].

Figure 11. Example of images from the LFW (Labeled Faces in the Wild) dataset.lfw.

4.8. CMU Multi PIE Dataset

The CMU Multi-PIE facial dataset [29] was developed between 2000 and 2009 at Carnegie Mellon University. It recovers more than 750,000 images of 337 persons acquired in up to four sessions under five months. Images were captured under 15 view-points and 19 lighting situations and presenting a variety of facial expressions. Also, high-quality frontal images were captured as well. In global, the dataset has more than 305 GB of facial data.

4.9. CASIA-WebFace Dataset

Another large-scale dataset for face recognition task, called CASIA-WebFace, was selected from the IMDb website with 10,575 persons and 494,414 facial images. It was built in 2014 by Yi et al. [30] at the Institute of Automation, Chinese Academy of Sciences (CASIA). It can be considered as an independent training-set for LFW. By this combination, the evaluation protocol of LFW can be standardized, and the reproducible research in face recognition in the wild can be advanced.

4.10. IARPA Janus Benchmark-A

The IARPA (Intelligence Advanced Research Projects Activity) Janus Benchmark A (IJB-A) was developed in 2015 by Klare et al. [31] to study face recognition/detection benchmarking. It contains videos and images in the wild from 500 subjects: 2085 videos and 5712 images with an average of 4.2 videos and 11.4 images per subject (Figure 12). To get a full pose variation and new situations than LFW, the facial images were recognized and restricted manually. The IJB-A established three protocols, two supporting both open-set identification and verification, and the third protocol is for detection. A separate face recognition protocol is presented. There are ten random training and

testing subclasses; for each subclass, 333 subjects are randomly placed in the training subclass, and the remaining 167 subjects are placed in a testing subclass. The search protocol uses probe templates for measuring the accuracy of the closed-set and open-set search on the gallery templates, and the protocol defines precisely which impostor and genuine comparisons must be performed for each subclass.

Figure 12. Examples of the faces in the IJB-A (Janus Benchmark A) database.

4.11. MegaFace Database

Shlizerman et al. [32] in 2016 introduced the MegaFace database, which includes 1,027,060 images of 690,572 different subjects. The MegaFace challenge uses a gallery to test the performance of face identification/verification algorithms with numerous "distractors", i.e., faces that are not in the test set, by training them from different probe set such as FG-NET [33] (includes 975 images of 82 persons with various ranges of ages) and FaceScrub [34] (includes 141,130 faces of 695 public figures).

4.12. CFP Dataset

CFP (celebrities in frontal-profile) is a public and challenging dataset, developed in 2016 by Sengupta et al. [35] at the University of Maryland. It includes 7000 pictures of 500 subjects (Figure 13). There are ten frontal images for each person, and more than four profile images. The evaluation protocol involves frontal-profile (FP) and frontal-frontal (FF) facial verification, each with ten folders of 350 pairs of the same person and 350 pairs of different persons.

Figure 13. Example images from the CFP (celebrities in frontal-profile) dataset.

4.13. Ms-Celeb-M1 Benchmark

Microsoft released the Ms-Celeb-M1 [36] large scale training benchmark in 2016, which contains around 10 million face images from 100k celebrities collected from the web to improve facial recognition technologies.

4.14. DMFD Database

To evaluate disguised face detection or recognition performance using disguised accessories, Wang et al. [37] created in 2016 the disguise covariate and/or make-up facial database with ground truth (goggle, beard, mustache, eye-glasses), acquired under real environments. This database contains 2460 images from 410 different subjects; most of these images are from celebrities (Figure 14). Three different protocols are considered: (1) protocol A calculates the corresponding scores on the all-to-all basis; (2) protocol B calculates the corresponding scores for one input image; and (3) protocol C uses the first images with the least obstruction (makeup, disguise, and wrong angle) for training while the rest of images are employed for testing.

Figure 14. Image pairs with a different type of makeup/disguise.

4.15. VGGFACE Database

VGGFACE (visual geometry group) is a large-scale training database assembled from the Internet by combining automation and humans in the loop. It contains 2.6 M images, over 2.6 K identities (Figure 15). It was created from the University of Oxford in 2016 by Parkhi et al. [38].

Figure 15. Example images from the VGGFACE (visual geometry group) dataset for six identities.

4.16. VGGFACE2 Database

In 2017, a large-scale face database called VGGFace2 was created by Cao et al. [39] from the University of Oxford. The database was collected from Google Images search with a wide range of age, pose, and ethnicity. It has 3.31 million images of 9131 identities, with 362.6 images for each identity on average (Figure 16). The VGGface2 is divided into two subclasses: the first is for training-set, including 8631 classes, and the second is for evaluation-set with 500 classes. Besides, two template annotations are described to allow assessment over pose and age: (1) Pose template: with five faces per template representing a consistent pose (frontal, profile, or three-quarter view) for 9 K facial images of 1.8 K templates; (2) age template: 400 templates (five faces per template with either an apparent age below 34, 34, or above) with 2 k facial images.

Figure 16. Example images from the VGGFACE2 dataset.

4.17. IARPA Janus Benchmark-B

The IARPA Janus Benchmark-B (IJB-B) database is an enlargement of IJB-A; it was created in 2017 by Whitelam et al. [40]. IJB-B consists of 1845 subjects for 21,798 still images (11,754 face and 10,044 non-face images) plus 55,026 frames from 7011 face videos. It is designed for detection, recognition, and clustering research in unconstrained environments. Different testing protocols were developed for IJB-B representing operational use cases, such as access point identification, surveillance video searches, forensic quality media searches, and clustering.

4.18. MF2 Dataset

The public dataset for face recognition MF2 (MegaFace 2) was created in 2017 by Nech and Shlizerman [41] of the University of Washington. It contains 672k persons and 4.7m images. MF2 was an attempt to create a benchmark to train algorithms on large-scale datasets and test at million scale *distractors* provided by the MegaFace challenge [36].

4.19. DFW Dataset

In 2018, Kushwaha et al. [42] created a novel disguised faces in the wild (DFW) dataset, consisting of 1000 subjects from 11,157 images with both obfuscated and impersonalized faces to improve the state-of-the-art for face recognition disguises. DFW defines three verification protocols, which are:

1. Impersonation protocol used only to evaluate the performance of impersonation techniques.
2. Obfuscation protocol used in the cases of disguises.
3. Overall performance protocol that is used to evaluate any algorithm on the complete dataset.

4.20. IARPA Janus Benchmark-C

The IARPA Janus Benchmark-C (IJB-C) database is an extension of IJB-B; it is developed in 2018 by Maze et al. [43]. It contains 31,334 still images (21,294 faces and 10,040 non-faces), with an average of 6 images per person, 117,542 frames from 11,779 full-motion videos, with an average of 33 frames and 3 videos per person. To advance state-of-the-art in unconstrained facial recognition, IJB-C defined different face detection protocols, 1:N identification (supporting closed-set and open-set evaluation), 1:1 verification, clustering, and end-to-end system evaluation which is a more operationally closed model of facial recognition use cases.

4.21. LFR Dataset

LFR (left-front-right) is a face recognition dataset presented by Elharrouss et al. [44] from Qatar University in 2020 to overcome pose-invariant facial recognition in the wild. Pose variation identifies a challenging facial recognition problem in an unconstrained environment. To deal with this issue, a CNN model for estimating pose is proposed. This model is trained using a self-collected dataset constructed from three standard datasets: LFW, CFP, and CASIA-WebFace, employing three classes of facial image capture: left, front, and right side. A dataset of 542 identities is thus generated, representing each subject's images on the left, front, and right face. Each folder of left and right includes 10–100 facial images, while the front folder contains 50–260 images.

4.22. RMFRD and SMFRD: Masqued Face Recognition Dataset

During COVID-19, nearly everyone wears a mask to restrict its spread, making conventional facial recognition technology inefficient. Hence, improving the recognition performance of the current facial recognition technology on masked faces is very important. For this reason, Wang et al. [45] proposed three types of masked face datasets, which are:

1. Masked face detection dataset (MFDD): it can be utilized to train a masked face detection model with precision.
2. Real-world masked face recognition dataset (RMFRD): it contains 5000 images of 525 persons wearing masks, and 90,000 pictures of the same 525 individuals without masks collected from the Internet (Figure 17).
3. Simulated masked face recognition dataset (SMFRD): in the meantime, the proposers utilized alternative means to place masks on the standard large-scale facial datasets, such as LFW [10] and CASIA WebFace [30] datasets, expanding thus the volume and variety of the masked facial recognition dataset. The SMFRD dataset covers 500,000 facial images of 10,000 persons, and it can be employed in practice alongside their original unmasked counterparts (Figure 18).

(**a**) (**b**) (**c**) (**d**)

Figure 17. Example images from the RMFRD (real-world masked face recognition dataset) dataset: (**a**,**b**) typical images, (**c**,**d**) masked images.

Figure 18. Example images from SMFRD (simulated masked face recognition dataset) dataset: simulated masked images.

Tables 1 and 2 resume and provide a comparative review of the cited above face recognition datasets. Table 1 contains datasets that can be used for training and/or testing, while Table 2 contains datasets that can be used only for training deep face recognition systems.

Table 1. Comparative summary of the most well-known/recent 2D face recognition datasets used for training and/or testing face recognition systems.

Database	Apparition's Date	Images	Subjects	Images/Subject
ORL [23]	1994	400	40	10
FERET [13]	1996	14,126	1199	-
AR [24]	1998	3016	116	26
XM2VTS [25]	1999	-	295	-
BANCA [26]	2003	-	208	-
FRGC [14]	2006	50,000	-	7
LFW [10]	2007	13,233	5749	≈2.3
CMU Multi PIE [29]	2009	>750,000	337	N/A
IJB-A [31]	2015	5712	500	≈11.4
CFP [35]	2016	7000	500	>14
DMFD [37]	2016	2460	410	6
IJB-B [40]	2017	21,798	1845	≈36.2
MF2 [41]	2017	4.7 M	672,057	≈7
DFW [42]	2018	11,157	1000	≈5.26
IJB-C [43]	2018	31,334	3531	≈6
LFR [44]	2020	30,000	542	10–260
RMFRD [45]	2020	95,000	525	-
SMFRD [45]	2020	500,000	10,000	-

Table 2. Comparative summary of the most well-known/recent 2D face recognition datasets used only for training deep face recognition systems.

Database	Apparition's Date	Images	Subjects	Images/Subject
CASIA WebFace [30]	2014	494,414	10,575	≈46.8
MegaFace [32]	2016	1,027,060	690,572	≈1.4
MS-Celeb-1M [36]	2016	10 M	100,000	100
VGGFACE [38]	2016	2.6 M	2622	1000
VGGFACE2 [39]	2017	3.31 M	9131	≈362.6

5. Two-Dimensional Face Recognition Approaches

A classical 2D face recognition system operates on images or videos obtained from surveillance systems, commercial/private cameras, CCTV, or similar everyday hardware. In a complete automatic configuration, the system must first detect the face in the input image/video and segment it from the detected area. Next, the must be aligned to some predefined canonical structure and treated to account for potential lighting changes. Features are extracted from the aligned/treated image, and last identity recognition is performed using a proper classification approach based on the calculated features.

Depending on the nature of the extraction and classification methods employed, we divide 2D face recognition methods into four different subclasses, namely: (1) holistic methods, (2) local (geometrical) methods, (3) local texture descriptors-based methods, and (4) deep learning-based methods, as illustrated in Figure 19.

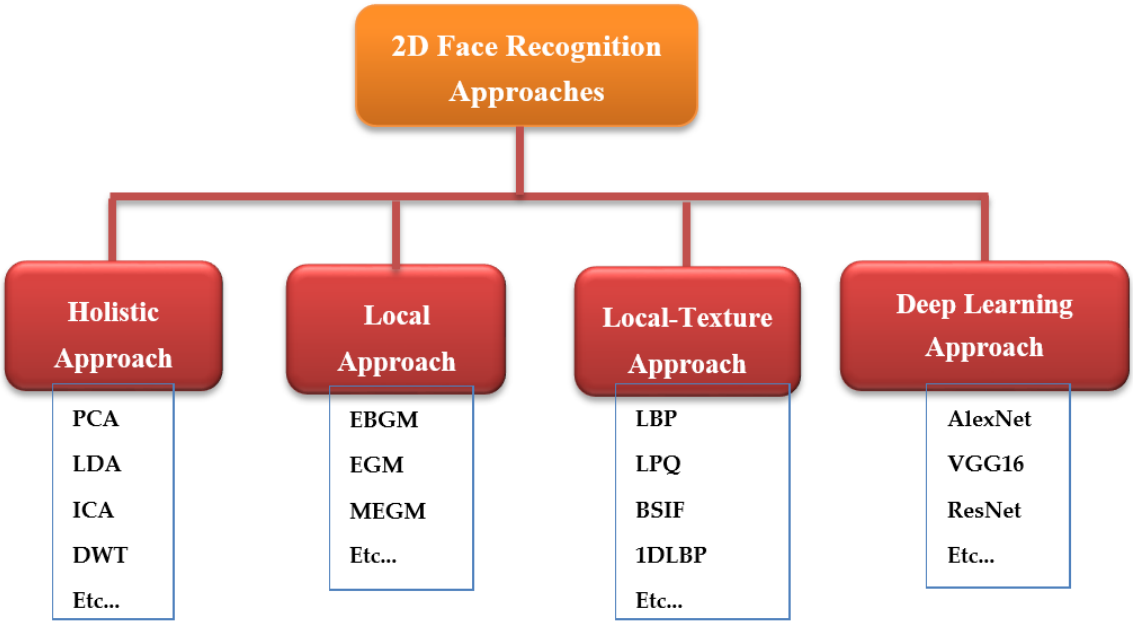


Figure 19. Taxonomy of 2D face recognition approaches presented in this paper.

5.1. Holistic Methods

Holistic or subspace-based algorithms assume that any M-collection of facial images holds redundancies that can be removed by applying the tensor's decomposition. These methods generate a collection of basis vectors representing a smaller space dimension (i.e., subspace) and preserving the original set of images. In the set of basis vectors, each face in the subspace can be reconstructed. To facilitate the operation, each facial image N × N is represented by a vector achieved by aligning the image rows. To find the non-singular basis vectors, the consequential matrix (N × N) × M is decomposed. Classification is frequently done by projecting a newly captured facial image and calculating the distance's measure with all classes described in that subspace. Besides, this approach's methods may be divided into two groups, namely linear and non-linear strategies, depending on representing the subspace. In this subsection, we only present some famous/well-known works in this approach because most of the published papers relevant to this subclass are too old.

Principal component analysis (PCA), known as eigenfaces [12], linear discriminative analysis (LDA), known as fisherfaces [46], and independent component analysis (ICA) [47] are the most common linear techniques employed for facial recognition systems.

In this approach, eigenface is considered as the pioneering and revolutionary method. It is also known as Karhunen-Loève expansion, principal component, or eigenvector. The works presented in the [48,49] references employed the PCA to efficiently characterize [50] the facial images. They have

shown that a few weights for each facial image and a standard facial image (eigenpicture) could approximately recreate any facial images. By projecting the facial image into the eigenpicture, the weights that model any face are attained.

Turk and Pentland [12] (1991) employed eigenfaces, influenced by Kirby and Sirovich's research [49], for face detection and recognition. Mathematically, the eigenfaces represent the facial distribution's main components or the eigenvectors of the facial image set covariance matrix. The eigenvectors are arranged respectively to model different quantities of the difference between the faces. So, a linear mixture of the eigenfaces can be precisely constituted for each face. It can also be calculated employing only the "best" eigenvectors with the greater eigenvalues. The top M eigenfaces build a space of M dimensions; the facial space. Using a private database containing 2500 images for 16 subjects, the authors reached 96%, 85%, and 64% on CCR (correct classification rate) under variations in illumination, orientation, and size.

To overcome the problem of performance degradation due to light variability, Zhao and Yang [51] (1999) presented a method for calculating the covariance matrix employing three images, acquired under different illumination conditions to account for random lighting effects when the subject is Lambertian.

Pentland et al. [52] (1994) extended their initial work from EigenFace to EigenFeatures concerning the facial elements, such as nose, eyes, and mouth. They employed a modular EigenSpace consisting of early EigenFeatures (i.e., EigenNose, EigenEyes, and EigenMouth). Compared with the original EigenFace method, this extended method showed less sensitivity to appearance variations. On the FERET dataset composed of 7562 images from about 3000 individuals, the authors achieved a CRR of 95%. EigenFace, compared to EigenFeatures, was a simple, fast, and practical method. Nevertheless, it does not present stability over changes in the conditions of illumination and scale.

Belhumeur et al. [46] (1997) suggested a system that would be insensitive to different lighting and facial expression changes. They regarded each pixel in the facial image as a point in a high-dimensional space. They observed that the image of a particular face resides in a 3D linear subspace of the high dimensional image space, in the condition where the front is a Lambertian area without resentment, in varying lights but under stable pose. If faces are not purely Lambertian areas and create self-umbrage, images will diverge from the linear subspace. Instead of directly stimulating this divergence, they projected the image linearly into a subspace to reduce specific areas of the face with a considerable alteration. The projection technique was based on LDA, which generated well-isolated classes in a small-dimensional subspace, even under various changes in lighting and facial expressions. The different experimental tests carried out on the Harvard and Yale Face databases showed that FisherFace has a lower error rate than EigenFace.

Barlett et al. [53] (2002) noted that PCA's baseline images are only dependent on pair-wise relationships between pixels in the image dataset. It seems appropriate to assume that superior basis images can be managed by methods sensitive to these high-arrange statistics in pattern recognition tasks, where vital information can be included in the high-arrange relationships between pixels. They employed the independent component analysis (ICA) [47], which is a PCA generalization. Besides, they implemented two different architectures with the FERET database to test ICA performance; the first process the pictures as random variables and the pixels as results, while the second processed the pixels as random variables, and the pictures as results. The first version defined for the faces spatially local basis pictures, and the second established a fractional facial code. The results of both ICA architectures under facial expression and aging were better than PCA. Besides, the best performance was achieved by fusing both ICA architectures.

Gabor filters (GFs) are spatial sinusoids positioned through a Gaussian window that enables images to obtain the characteristics by choosing only their frequency, orientation, and size. Abhishree et al. [54] (2015) suggested a method based on GFs to extract features and enhance the performance of face recognition systems. GFs are employed for capturing aligned facial characteristics at specific angles. Besides, an optimization technique of feature selection is employed to find the optimal feature space.

The method proposed was tested under multiple databases, such as ORL and FERET, and showed good results against variations in posture, lighting, and expression variations.

Discrete cosine transform (DCT) [55] and discrete wavelet transform (DWT) [56,57] are other linear-techniques that have been employed for facial analysis. Both methods are employed mainly in image compression [58] and feature selection. Wang et al. [59] (2010) suggested a DWT and DCT-based fused feature extraction algorithm for face recognition. The face's image is decomposed using 2D-DWT, and then 2D-DCT is employed to approximate the low-frequency image received from the preceding step. Finally, the DCT coefficients are employed for matching. The experimental results of the ORL database showed the superiority of this algorithm compared to the traditional PCA.

To sum up, all holistic methods are prevalent in the implementation of face recognition systems. Nonetheless, they are very pone to context changes and misalignments. For this reason, the face must be cut manually from the image in the majority of cases. On the other hand, as the data set is viewed as a single matrix, it is necessary to enforce geometric consistency in all facial instances. Thus, all facial images must be carefully matched within a standard frame of reference. A minor error in face orientation can cause substantial facial classification errors.

5.2. Geometric Approach

Attention and fixations play a crucial function in human face recognition. Attentive processes are usually guided by landmark characteristics localized in the considered space by calculating a salience map. The same landmarks may offer useful information when faced with algorithms for recognition. The facial regions in the image do not provide the same amount of information. The forehead and cheeks, for example, have straightforward structures and fewer distinctive patterns as compared to the nose or eyes. The landmarks in the face are used to register facial features, the normalization of expressions, and the recognition of defined positions based on the geometric distribution and the grey level pattern. Even though the in-depth studies summarized in craniology accurately represent a rich set of facial landmarks for face recognition, there is no universal set of admitted landmarks.

Bookstein [60] (1989) described the landmarks as: "points in a form for which biological counterparts that occur in a collection of data, which are objectively significant and reproducible, in all other forms." The most commonly employed landmarks on the face are the tip of the nose, the tips of the eyes, the tips at the corners of the mouth, the eyebrows, the middle of the iris, the top of the ear, the nostrils, and the nasal. It should be noted that discriminating regions of the face, such as eyes or mouth, are also called "facial features" in the literature. Sometimes that terminology leads to ambiguity. Indeed, in pattern recognition, the term "feature" is most often employed to specify a particular representation extracted from a pattern at the grey level. For instance, the EigenFace vectors are also called "features." Furthermore, the numerical representations collected by the multi-channel Gabor implemented to a grey-level picture are named "features." For this intention, patterns derived from particular and discriminating positions are here referred to as "landmarks" rather than "features."

The distribution of landmarks is employed in geometric-based methods in the structure of heuristic rules involving distances, angles, and regions [61,62]. Geometry is organized into a full model of building in structure-based methods. For example, in the elastic bunch graph-matching (EBGM) algorithm [63], a graph models the positions related to the landmarks, wherever every node denotes one point on the face, and the arcs are weighted in accordance to the mark's predictable distances, as shown in Figure 20. A series of models are employed to determine the similarity of the local characteristic for each node. Although the possible deformations often depend on landmarks (e.g., the mouth corners deform much more than the nose tip), the specific landmark information can be joined toward the structural model [64]. With the expansion of the set of jointly optimized constraints, the system works more often with problems of convergence and local optimums, which in effect necessitates a successful- and sometimes manual-initialization.

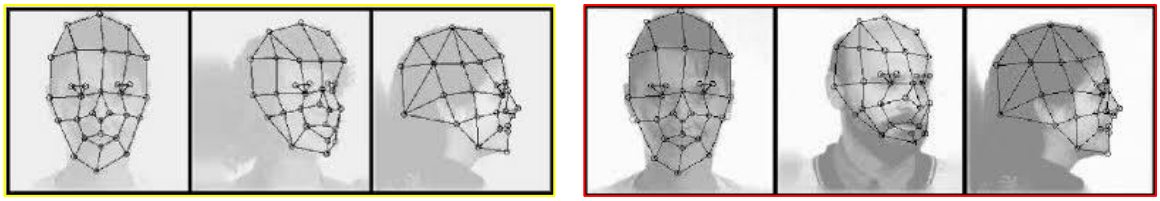


Figure 20. Example of extraction of landmarks using the elastic bunch graph-matching (EBGM) algorithm.

Several methods were proposed to derive facial representations from several components or sub-images of the face. Pentland et al. [52] (1994) suggested a PCA version based on components whose facial subspace consisted of some subspaces constructed from partial pictures of the initial facial images. The selected landmarks were caught between the mouth and the eyes.

Tistarelli [65] (1995) suggested a system focused on extracting facial references re-sampled by a log-polar mapping program. The identification was carried out through the application of cross-correlation and normalization between two facial designs. The correlation value determined the resemblance between the two images, and consequently, the two subjects. For the classification, the correlation value was employed as a ranking.

One common form of face recognition based on landmarks is elastic graph matching (EGM) [66,67]. EGM is a realistic implementation of dynamic arc construction for object identification. The referential object graph was generated with EGM by superimposing a sparse, elastic, and rectangular graph on the object's image and determining the Gabor wavelet bank's response to every node of the graph. The cumulative value represents one jet at each node. Stochastic optimization of a loss function that considers the similitudes of the jets and the node deformation introduced the process of graph matching. A two-step optimization is necessary to minimize this loss function. Lades et al. [66] (1993) reported exciting results on a private dataset of 87 persons under different facial expression variations and 15° rotation.

The elastic bunch graph-matching (EBGM) [68,69] is an expansion of EGM. In the heap graph structure, a collection of jets was calculated for various examples of the same face at each node (e.g., with the open or closed mouth and eyes). In this form, the heap graph representation can handle several changes in facial appearance.

A further technique close to EGM is morphological elastic graph matching (MEGM) [70,71]. The Gabor characteristics are substituted by multi-scale morphological characteristics achieved by filtering facial image with dilation-erosion.

Kumar et al. [72] (2020) proposed an ensemble face recognition system that employed a novel descriptor named dense local graph structure (D-LGS). Besides, the descriptor employs a bilinear interpolation to improve the pixel density when generating the graphic picture from the entered image. It did well on both constrained (e.g., ORL database) and unconstrained (e.g., LFW database) environments.

In summary, the major disadvantage of all methods based on geometry is that they involve perfectly aligned facial images. All facial images must be aligned to possess all referential points (e.g., mouth, nose, and eyes) displayed at the corresponding place's feature vector. The facial images are most often manually arranged for this purpose, and are often put under an anisotropic scale; the optimal automatic alignment is usually considered a challenging task. By comparison, EGM does not need precise alignment to work well. The EGM's critical drawback is the time taken to examine the facial image in different scales and the matching technique. It is generally known that in this perspective, the variations in lighting that contemplate face recognition present one of the significant challenges. How computers design human face geometry is considered another issue that researchers are invited upon to resolve to improve the robustness and safety of face recognition systems.

5.3. Local-Texture Approach

Feature extraction strategies focused on knowledge about the texture play a significant role in pattern recognition and computer vision. Texture extraction algorithms suggested in the literature can be subdivided into statistical and structural methods [73–77]. Local texture descriptors subsequently gained more attention and were introduced in many applications, such as texture classification, face recognition, or image indexing. They are distinctive, resilient to monotonic gray-scale changes, poor lighting, variance in brightness, and do not need segmentation [78]. The local descriptor's goal is to transform the information at pixel-level into an appropriate form, which acquires the most compelling content insensitive to different aspects induced by variations in the environment. Contrary to global descriptors that calculate features directly from the entire image, local descriptors, which are more efficient under unconstrained situations, model the elements in small local image patches [79].

Ahonen et al. [80,81] (2004–2006) outlined the groundbreaking work of this approach. The authors presented a novel and effective representation of the facial image based on the local texture descriptor named: local binary pattern (LBP). The facial image was separated into different blocks, where the distributions of the LBP feature were selected and combined into an improved histogram used as a facial descriptor, as shown in Figure 21. The texture representation of a single area encodes the area's appearance, and the combination of descriptions of the entire area defines the face's global morphology. The Colorado State University Face Identification Evaluation System protocol [82] and the FERET database were used to measure the face recognition issue's performance. The works related to this were used for comparison: Bayesian intra/extra-personal classifier (BIC) [83], PCA [12], and EBGM [63]. The first findings of this study were published in the 2004 ECCV conference [80]. Besides, the authors presented an in-depth analysis of the suggested method in [81]. The weighted LBP (e.g., CRR = 97% with *Fb* probe-set) yielded higher recognition rates than other similar works (CRR = 86% with PCA and 89% with EBGM under the same probe-set). The LBP showed robustness to many difficulties provoked by illumination variations or aging of the person, but more research was still required to achieve even more reliable performance.

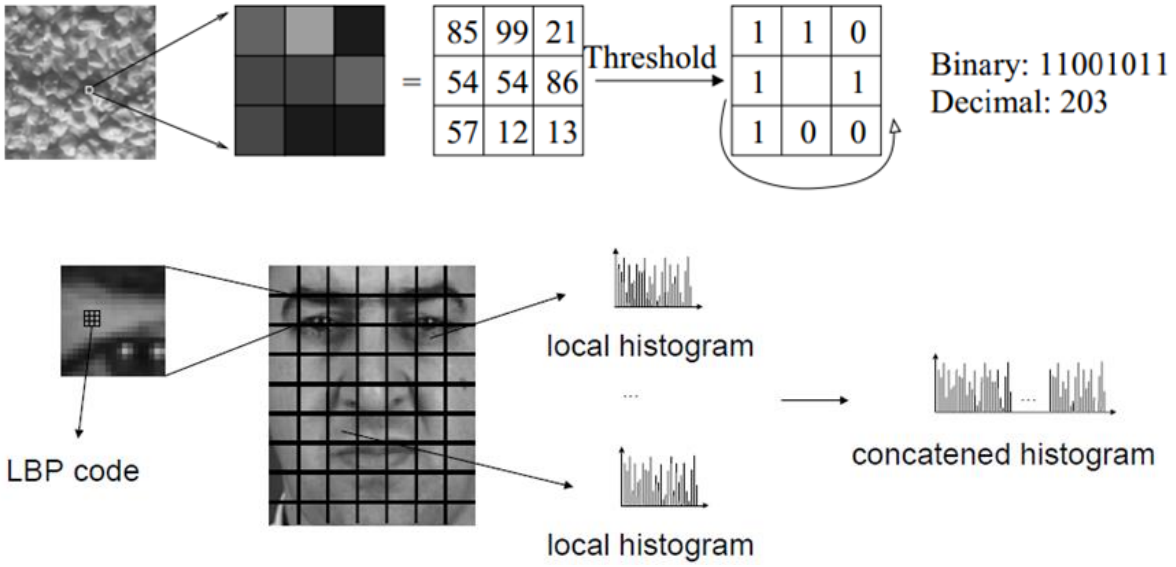


Figure 21. Example of facial local binary pattern (LBP) calculation.

Rodriguez and Marcel [84] (2006) suggested a generative method to face verification based on LBP facial representation as a complementary work. They created a universal facial model as a series of LBP-histograms; the histograms extracted from each block were seen as a distribution probability rather than a statistical observation. Next, by applying the Maximum A Posteriori (MAP) adaptation technique to the generic model under a probabilistic system, they obtained a client-specific model.

Two primary facial verification datasets, namely XM2VTS and BANCA, were employed to assess the suggested method's performance. The HTER (half total error rate) obtained with LBP/MAP was 1.42% using the XM2VTS database; LBP/MAP performed better compared to two other LBP related works. With the BANCA database, they noticed that the proposed method performed better under all conditions than two other LBP-based methods. The LDA/NC method [85] reported the best results under matched conditions with HTER = 4.9%, while the alternative solution ranked 3rd with HTER = 7.3%. Nevertheless, the LBP/MAP outperformed the LBP/JSBoost method [86] when more substantial training data are available. LBP/MAP showed the best results for the unconstrained environment, particularly in degrading illumination.

Following the previous work, Boutella et al. [87] (2014) noted that the histograms extracted from blocks are mostly sparse in the original LBP facial representation and its variants; most bins in the histogram are near to zero or zero, especially in short blocks. Additionally, large blocks provide dense histograms that are not effective in representing local facial changes. The authors applied a vector quantization (VQ) on each block to get a useful feature vector; i.e., each block's patterns are grouped into several groups, and the face is described by a codebook containing only valid LBP labels and ignoring other inefficient labels. They also developed a reliable face model through adaptation to MAP. The databases XM2VTS and BANCA were used to measure the performance of this method named: VQ-MAP. The obtained results (e.g., HTER = 0.8% with XM2VTS) showed promising results which exceeded the original LBP results (Ahonen et al. [81], Rodriguez et al. [84], and several related works). In Mc (match controlled) and Ua (unmatched adverse) protocols, the proposed solution produced competitive results for the BANCA database. Average performances are obtained for the remaining protocols (Ud (unmatched degraded), P (pooled test), and G (grand test)). The suggested solution was characterized by the simplicity and computational effectiveness of the baseline LBP, as opposed to the comparative methods.

Inspired by the original LBP, Benzaoui and Boukrouche [88–90] (2013–2014) proposed a new representation of the LBP operator, projected in one-dimensional space, called: one-dimensional local binary pattern (1DLBP), to recognize faces. As shown in Figure 22, they decomposed the feature extraction algorithm into five main steps; first, the image entered was decomposed into several blocks of the same size. A vertical projection was applied to each decomposed block, in one-dimensional space. Furthermore, the proposed 1DLBP descriptor was applied to every projected block. Then, they concatenated the vectors generated from each block to create one global vector. Finally, the principal component analysis (PCA) was used to regroup the global vectors, reduce the dimensionalities, and keep only each individual's relevant information. Chi-square distance was used to calculate the similarity between the images of the face. They performed several experiments on the ORL and AR datasets; they found that the projected 1DLBP operator (e.g., CRR = 96.9% on AR the database) was very successful than 2D LBP (86.4%). The authors also expanded their work by adding the K-NN algorithm classification and the combination of vertical and horizontal vectors projected from each block.

Ahonen et al. [91] (2008) used the newly introduced operator: local phase quantization (LPQ) [92] in recognition of blurred faces to solve the facial recognition issue under blurring situations. The LPQ operator is based on the quantization in local regions of the Fourier transform phase. The phase was considered a blur invariant property following specific conditions that were frequently met. In their proposition, LPQ label histograms calculated in local neighborhoods were employed as a facial descriptor in the same way as the commonly used facial description LBP methodology. Datasets of CMU PIE and FRGC were used in the various tests. The experiments on the CMU PIE dataset with synthetically induced Gaussian blur in the probe-set showed that the LPQ descriptor (98.6% with max standard deviation $\sigma = 2.0$) is very robust to blur compared to LBP (93.5%). Furthermore, its efficiency was superior to LBP (92.7%), even with no blur (99.2%). The performance of the LPQ operator (74.5%) outperformed all comparative methods, LBP (64.3%) and LTP [93] (68.4%), for the FRGC dataset where probe-images contain several lighting, facial expression, and blur variations. The authors deduced that LPQ is very useful at blurring and in unconstrained conditions, such as changes in facial expression

and lighting. As an advantage, the operator is simple to compute and easy to implement, requiring just image convolutions with little independent kernels and rotations of vectors.

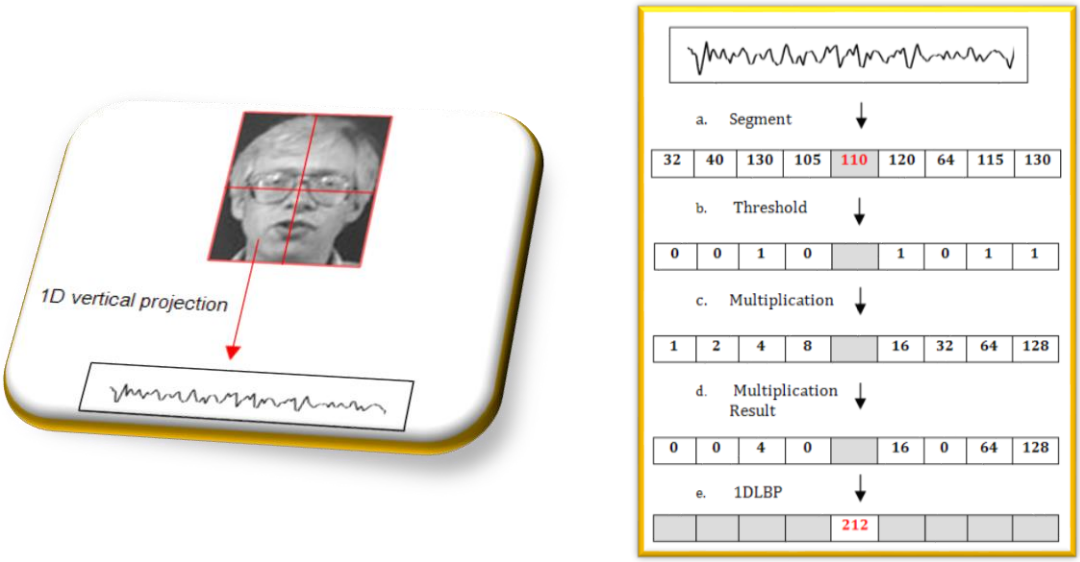


Figure 22. Example of one-dimensional local binary pattern (1DLBP) facial calculation.

The recognition of faces from low-resolution images is considered a challenging problem. Lei et al. [94] (2011) proposed an adequate local frequency descriptor (LFD) to surmount this issue. Like LPQ, the proposed descriptor is based on local frequency information, making it robust to low-resolution and blur. Different from LPQ, the descriptor employs both phase and magnitude information, thus providing more information. Furthermore, the LFD was determined not to allow the positive PSF (point spread function) assumption for the blur kernel. Also, the representation is expected to be very useful because more information was exploited in the frequency band's border. Besides, they implemented a statistically uniform scheme of pattern interpretation to improve the method's effectiveness. The proposed LFD descriptor's performance was compared with two related descriptors, LBP and LPQ, using facial images of low-resolution provided by the FERET database. Two experiments were designed to produce specific images of low-resolution. For the first one, the probe images were adjusted into 88×80, 66×60, 44×40 and 33×30. In the second experiment, the motion blur problem was simulated in the probe-set by adding the shift-invariant linear blur PSF. The production of LPQ and LFD in Fb set (right conditions) is very similar (the LPQ was the highest in all cases). The performance of LFD significantly outperformed LBP and LPQ in FC (illumination variation), dup-1, and dup-2 probe sets. The LFD's high-quality performance with various low-resolution images showed that LFD is robust and useful for real-world applications.

Kannala and Rahtu [95] (2012) proposed a method for building local image descriptors that encode texture information efficiently and are proper for the description of image regions based on histograms. Based on LBP and LPQ, the suggestion behind the proposed method called BSIF (binarized statistical image features) is to train a fixed collection of filters from a limited number of original images automatically as an alternative to employing hand-crafted filters such as LBP and LPQ. To get a statistically significant form of the images, BSIF uses learning as a substitute for manual tuning, allowing specific information encoding employing uncomplicated element-wise quantization. Learning also offers a comfortable and versatile method for adapting the descriptor's size and controlling it to applications with abnormal image properties. They applied the FRGC in the experiments; the BSIF method results were in a similar performance to the state-of-the-art methods (≈75%). However, some of the probe-images were imperfectly aligned and blurred; BSIF ranked comparable output to explicitly developed methods of rotation and blur invariant. The authors showed the BSIF method's tolerance to image degradations commonly found in real-applications.

In summary, this approach's methods are characterized by the advantage of high efficiency in time analysis and the rate of recognition. They are easy to incorporate, which allows examining photographs in real-time in a demanding environment. Besides, they are invariant to scale and misalignment. However, they are characterized by the complexity of automatic detection of relevant features and the inability to discriminate. They also suffer in the following situations: variations of posture, low resolution, facial expression, and different illumination conditions.

5.4. Deep Learning Approach

5.4.1. Introduction to Deep Learning

Deep artificial neural networks, known as deep learning, have won copious contests in the pattern machine learning and recognition over the past few years [96]. Deep learning, belonging to a machine learning class, employs successive hidden-layers of information-processing levels, hierarchically organized for representation or pattern classification, and feature learning [97]. According to Deng and Yu [98] (2014), there are three principal reasons for the prominence of deep learning: starting with the drastic growth of processing abilities (e.g., GPU units), and the dramatically lower computing hardware costs, and finally the recent progress in machine learning studies. Many researchers proved successful deep learning results in diverse applications of computer vision, conversational speech recognition, phonetic recognition, voice search, speech and image feature coding, hand-writing recognition, semantic utterance classification, visual object recognition, and audio processing, and information retrieval [97]. Deep learning can be categorized into three main classes depending on how the technique and architecture are used:

1. Unsupervised or generative (auto encoder (AE) [99], Boltzman machine (BM) [100], recurrent neural network (RNN) [101], and sum-product network (SPN) [102]);
2. Supervised or discriminative (convolutional neural network (CNN));
3. Hybrid (deep neural network (DNN) [97,103]).

Discriminative deep architectures or supervised learning are supposed to differentiate several parts of data for classification. CNN is the best example of supervised learning; it allows exceptional architectural proficiency for image recognition [103]. Face recognition is commonly studied in computer vision, and CNN has achieved great success, becoming a powerhouse in this topic. This sub-section focuses on how the so-called powerhouse was used in its full effectiveness in face recognition.

5.4.2. Convolutional Neural Networks (CNNs)

CNN's are a form of neural networks that have proved successful in areas such as the recognition and classification of images. CNNs consist of a set of filters/kernels/neurons with learnable parameters or weights and biases which have been added. Each filter takes some inputs, makes convolution, and follows it with a non-linearity. The structure of CNN includes layers of convolutional, pooling, rectified linear unit, and fully connected.

- Convolutional layer: This is the CNN's core building block that aims at extracting features from the input data. Each layer uses a convolution operation to obtain a feature map. After that, the activation or feature maps are fed to the next layer as input data [9].
- Pooling layer: This is a non-linear down-sampling [104,105] form that reduces the dimensionality of the feature map but still has the crucial information. There are various non-linear pooling functions in which max-pooling is the most efficient and superior to sub-sampling [106].
- Rectified linear unit (ReLU) Layer: This is a non-linear operation, involving units that use the rectifier.
- Fully connected layer (FC): The high-level reasoning in the neural network is done via fully connected layers after applying various convolutional layers and max-pooling layers [107].

5.4.3. Popular CNN Architectures

LeNet

LeNet refers to LeNet-5, proposed in 1998 by Lecun et al. [108]. It is a capable CNN trained with the back-propagation algorithm for handwriting digit recognition. LeNet-5 consists of seven trainable layers, two convolutional, two pooling, and three fully connected layers. LeNet is regarded as the backbone of modern CNN.

AlexNet

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is a benchmark for large scale of object recognition in annual competitions starting from 2010 to present [109]. Krizhevesky et al. [110] have won ILSVRC-2012 using a large deep CNN model, named AlexNet, that have achieved record-breaking results in computer vision approached against all the traditional machine learning. AlexNet comprises five convolutional layers, some of which are followed by max-pooling, and three fully connected layers with 1000 way softmax, as shown in Figure 23, and other techniques, such as dropout, rectified linear unit (ReLU), and data augmentation.

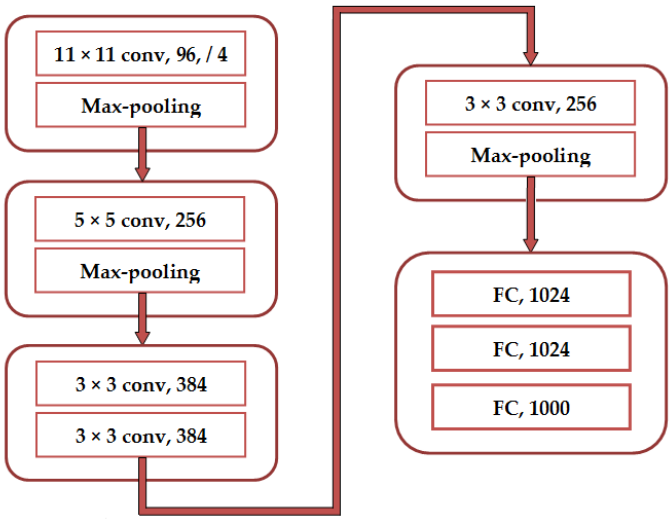


Figure 23. AlexNet architecture. FC (fully connected layer): fully connected layers. conv: convolution.

VGGNet

At ILSVRC-2014, Simonyan et al. [111] have explored how convolutional network depth affects the accuracy of the image recognition setting on a large scale. Their principal contribution was to use an architecture called VGGNet with small (3 × 3) convolution filters and double the number of feature maps after the (2 × 2) pooling. The network's depth was increased to 16–19 weight layers, improving the deep architecture flexibility to learn continuous nonlinear mappings, as shown in Figure 24.

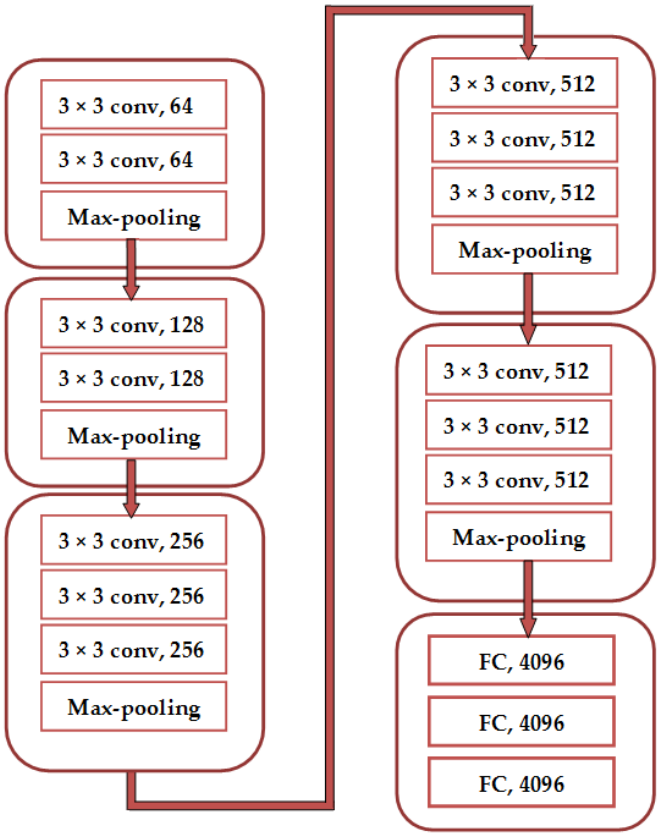


Figure 24. VGGNet architecture.

GoogleNet

The winner of ILSVRC-2014 was the 22-layer GoogleNet, a model proposed by Szegedy et al. [112] (2014), to minimize computational complexity compared to the standard CNN model. It introduced an "inception module," containing variable receptive fields generated by different kernel sizes. Several convolutions (1 × 1, 3 × 3, and 5 × 5) and (3 × 3) max-pooling are effectuated in parallel for the previous input and output. All feature maps are also concatenated together as the input of the next module, as shown in Figure 25.

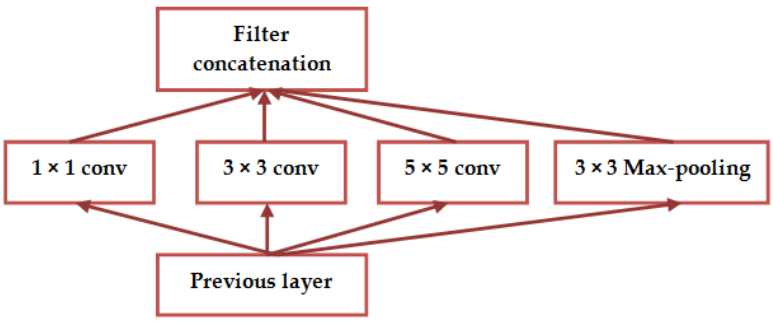


Figure 25. GoogleNet architecture.

ResNet

He et al. [113] (2015) introduced a novel architecture named residual neural network (ResNet) to facilitate the training of ultra-deep networks compared to networks already in use. ResNet was the winner of ILSVRC 2015; it was developed with "shortcut connections" and features batch normalization, it was able to train a neural network with various numbers of layers: 34, 50, 101, 152, and even 1202. Figure 26 illustrates the basic block diagram of the ResNet architecture.

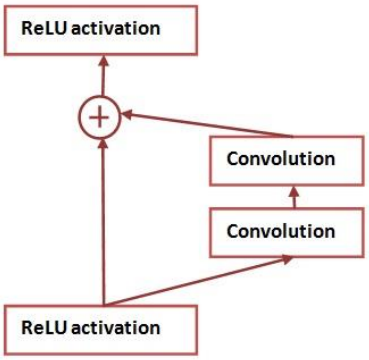


Figure 26. ResNet architecture.

SENet

Hu et al. [114] won the first place at ILSVRC-2017 since they proposed the block squeeze-and-excitation (SE), a novel architecture unit, which recalibrates channel-wise feature responses by clearly modeling the inter-dependencies between channels. The SE network (SENet) was developed by stacking a set of SE blocks and can be integrated with standard architecture such as ResNet, improving their effectiveness in numerous datasets and tasks (Figure 27).

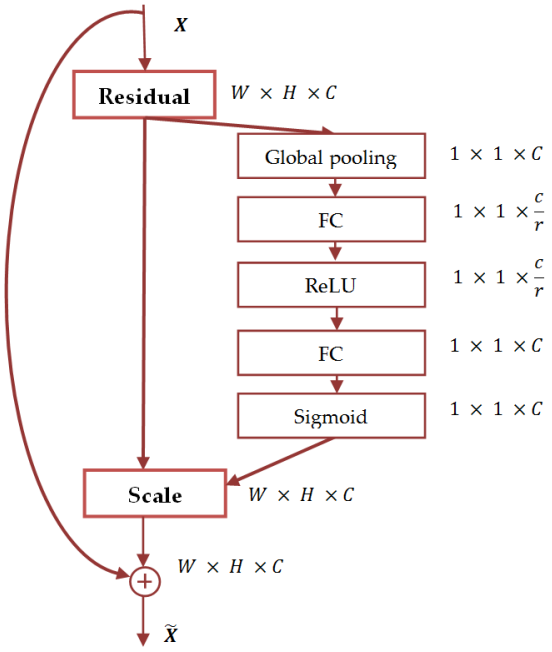


Figure 27. SENet architecture. X: the input. $W \times H$: spatial dimensions. C: channel descriptor. r: reduction ratio. FC: fully connected layers. $\hat{X}$: the output.

5.4.4. Deep CNN-Based Methods for Face Recognition.

In the following, we discuss several deep face recognition methods based on CNN, which are typically trained in a supervised manner. There are many significant lines of research to train deep CNNs for face recognition. Those that train a multi-class classifier separate diverse facial identities in the training phase, such as using the Softmax classifier; those that are trained to learn more discriminative in-depth face features. Some methods focus on extracting features from various facial regions using multi-CNNs or focus on extracting appearance variations features from non-frontal facial images. Many works adopted ideas from metric learning and combined different loss functions or using effective loss function methods, and others used proper activation function. The following works are organized according to their architecture.

Investigations Based on AlexNet Architecture

DeepFace, proposed by Taigman et al. [15] (2014), is a multi-stage approach that uses a generic 3D shape model to align faces. They have derived a facial representation from a 9-layer deep neural network, trained from a multi-class face recognition task on over 4000 identities. The authors also experimented with a Siamese network [115] in which the top-layer distance between two facial features was directly optimized. DeepFace was one of the first works that achieved very high accuracy on the LFW dataset using CNNs. Inspired by this study, the focus of face recognition research has moved to deep-learning-based approaches, in just three years, the accuracy was dramatically increased.

DeepFace has been expanded to other works like the DeepID series presented in several papers by Sun et al. [116–119], where they steadily increased the performance on LFW. In [116], they suggested the Deep hidden IDentity features (DeepID) to learn the verification task's high-level face feature representations. The features are obtained from each deep convolutional network's final hidden layer and predict around 10,000 identity classes in the training set. The number of features keeps on lessening along the feature extraction hierarchy until the DeepID layer. Those features are extracted from different facial regions to shape over-complete and complementary representations and further improve the LFW performance with just weakly aligned faces compared to DeepFace [15].

One of the main challenges of face recognition is to develop an efficient feature representation for reducing intra-personal variations while increasing inter-personal variations, which can be solved with the Deep IDentification verification features (DeepID2) [117]. The features were learned with variations of deep CNNs under two supervisory signals: (1) the signal of face identification raises the inter-personal variations by drawing DeepID2 derived from separate identities apart. (2) The signal of face verification decreases the intra-personal by pulling together DeepID2 derived from the same identity. The merging of these two supervisory signals results in far better features than either.

The authors presented in [118] a high-performance deep convolutional neural network (DeepID2+) for face recognition that has improved upon the DeepID2 [117], by augmenting the dimension of hidden representations and joining supervision to early convolutional layers.

Taigman et al. [120] (2015) have applied a Semantic bootstrapping method to replace the naive random sub-sampling [121] of the training set for selecting from an extensive database an efficient training-set. Besides, they have also discussed further stable protocols of the LFW dataset, indicating a robust representation of the CNN facial features.

Liu et al. [122] (2015) have proposed a two-stage approach for face recognition, which extracts low dimensional yet high discriminative features by merging a multi-patch deep CNN with deep metric learning. They have found local patches that are less sensitive to variation, mainly expressions, and poses.

Investigations Based on VGGNet Architecture

In [119], Sun et al. proposed two deeper neural network architectures for face recognition referred to as DeepID3, deeper than DeepID2+. These two architectures are reconstructed from crucial elements of GoogLeNet [112] and VGGNet [111] that are: stacked convolution and inception layers. Joint facial identification-verification supervisory signals were added during training-set to both final feature extraction and intermediate layers.

Parkhi et al. [38] (2015) have made two contributions: first, they have developed a method for assembling a large scale dataset (VGGface dataset), with little label noise. Second, the authors have investigated diverse CNN architectures for face recognition based on VGGNet [111], including exploring a triplet loss function and face alignment. To allow a direct comparison with previous work, they have applied their proposed dataset, VGGFace, for training, and the evaluation was performed on the famous benchmark database, LFW.

Data augmentation aims to increase the dataset by making transformations on the images without changing the labels, which have been commonly utilized to improve the CNN performance and prevent overfitting. Masi et al. [123] (2016) have augmented their training data by generating new

facial images with specific appearance variations, including shape, pose, and expression using a 3D generic face.

To investigate the long tail effect in deep facial recognition, Zhang et al. [124] (2016) have proposed a new function named Range Loss that uses the harmonic mean value to reduce intrapersonal variations yet enlarge inter-personal differences.

Liu et al. [125] (2016) have proposed a generalized Large-Margin Softmax Loss (L-Softmax), which combines the most generally used components in deep CNN architectures, that are: a cross-entropy loss, a Softmax loss, and the final fully connected layer. The L-Softmax loss defines a flexible learning objective with the adjustable margin and can avoid overfitting. The experimental results on different datasets have shown that L-Softmax loss boosts performance in verification and classification tasks.

Chen et al. [126] (2017) proposed a Noisy Softmax to alleviate early individual saturation problems by injecting annealed noise in the Softmax, which aims to improve CNNs generalization capacity.

Investigations Based on GoogleNet Architecture

FaceNet is a model from Google proposed by Schroff et al. [127] (2015) that uses 128-dimensional representations from deep convolutional networks, trained on 200-million facial images by utilizing a triplet loss function at the final layer. The triplet consists of two matching facial patches and a non-matching facial patch, and the loss attended to separate by a distance margin the positive from the negative pair. This loss is further suitable for face verification. They discussed two various core architectures: NN1 based on the Zeiler and Fergus model networks [128] and NN2 based on the style inception networks [112] from GoogLeNet.

Ben Fredj et al. [129] (2020) used aggressive data augmentation with randomly perturbing information and complicated facial appearance conditions. One of the main ideas was to use the adaptive fusion strategy of softmax loss and center loss, improving performance, and making the model more flexible and efficient.

Investigations Based on LeNet Architecture

Wen et al. [130] (2016) were the pioneers to introducing a supervisory signal, namely center loss, for face recognition research, which learns for each class a center for deep face features while simultaneously reducing the distances between the features and matching class centers. Thus, the learned face features' discriminative power is enhanced, and variations in the intra-class feature are minimized.

Wu et al. [131] (2017) proposed a center invariant loss function, which aligns each individual's center to enforce the deeply learned facial features to have more general representations for the entire people. Thus, better separation of feature space for all classes gives highly imbalanced training data.

Yin et al. [132] (2019) have introduced another work that tried to solve the imbalance of training data. They have proposed a novel feature transfer learning (FTL) that adapts under-represented (UR) classes' feature distribution, resulting in training less biased deep face recognition.

Investigations Based on ResNet Architecture

A related motivation to the feature normalization was proposed by Ranjan et al. [133] (2017), which have used an L2-constraint on the Softmax loss for training a facial verification system. The L2-Softmax loss realizes compact feature learning by constraining the deep features to lie on a given radius's unit hypersphere.

To develop the discriminative power of the deep features, Deng et al. [134] (2017) have proposed the marginal loss function, which increases the inter-class separations and reduces the intra-class variations with the joint supervision of Softmax and marginal loss.

The NormFace loss was proposed by Wang et al. [135] (2017) for improving the task of face verification. It studies and identifies the issue of applying L2 normalization operations on the embeddings and the weight vectors of the output layer before Softmax. Two training strategies for

normalized features are proposed: the first is a reformulated Softmax loss by replacing the inner-product with cosine similarity. The second is inspired by metric learning.

Liu et al. [136] (2017) proposed a metric learning loss that is congenerous cosine (COCO) for the individual recognition task. Their idea consists of optimizing and comparing the cosine distance between deep features to be polymerized and discriminated. COCO loss is expected to have lesser maximal intra-class variation than minimal inter-class distance.

Hasnat et al. [137] (2017) proposed to model deep-feature learning from deep CNN as a mixture of von Mises-Fisher distributions, by integrating von Mises-Fisher (vMF) mixture models with deep CNN model. They derived a novel loss function called *vMFML* that allows for discriminative learning.

Liu et al. [138] (2018) have presented a Deep Hypersphere Embedding method for face recognition (SphereFace). In particular, they proposed an angular Softmax loss (A-Softmax), which allows deep CNN to learn discriminative facial features with the angular margin by imposing constraints on a hypersphere manifold.

Zheng et al. [139] (2018) introduced a feature normalization approach for deep CNN, called ring loss, to normalize all samples of facial features via convex augmentation of the standard loss function (like Softmax). Ring loss applies soft feature normalization, where it ultimately learns to constrain facial feature vectors on the unit hypersphere.

To tackle the imbalance of training data, Guo and Zhang [140] (2018) established a multiclass classifier by using Multinomial logistic regression learning (MLR). MLR trains the Softmax classifier in combination with the underrepresented classes promotion (UP) loss term. They called this term as classification vector-centered Cosine similarity (CCS) loss, which improves one-shot face recognition accuracy.

Wang et al. [141] (2018) have proposed a novel loss function, called large margin Cosine loss (LMCL) to conduct deep CNNs to learn more discriminative features for face recognition. They reformulated the Softmax loss as a cosine loss by L2 normalizing weights and feature vectors to eliminate radial variations. Accordingly, to optimize the decision margin in angular space, a cosine margin concept is introduced, maximum inter-class variance and minimum intra-class variance are realized. Based on LMCL, the authors have constructed a deep model, namely CosFace. In the training set, the discriminative features are learned with a large cosine margin, and the features are extracted from deep CNNs in the test set to perform either face identification or face verification.

Wang et al. [142] (2018) have proposed to impose a novel additive margin intended for the Softmax loss. The margin was formulated via a cosine similarity with normalized weights and features, resulting in improved learning face representations.

Wu et al. [143] (2018) have presented a light CNN framework that works on the massive datasets with noisy labels. The authors first introduced max-feature-map (MFM), a variation of maxout activation that uses a competitive relationship. They have also introduced three networks that reduce the computational costs and number of parameters. They have finished their work by presenting a semantic bootstrapping that predicts which network is more consistent with noisy labels.

To tackle class-imbalanced learning using deep CNN, Hayat et al. [144] (2019) have proposed the first hybrid loss function based on an affinity measure in Euclid space that aims at realizing a generalizable large margin classifier. The proposed loss combines clustering and classification in a single formulation that reduces intra-class variations while concurrently achieving maximal inter-class distances. Experimental evaluations have shown the effectiveness of the affinity loss function for face verification using several datasets that present a natural imbalance.

In order to ameliorate the discriminative power of deep CNN features for face recognition, Deng et al. [145] (2019) have proposed an additive angular margin loss (ArcFace), which has an optimized geometric interpretation that improves the geodesic distance margin by matching the arc and the angle in the normalized hypersphere.

Many data in face analysis tasks, including face recognition and face-attribute prediction, can naturally exhibit imbalanced class distribution, i.e., most data belong to some majority

classes. In contrast, minority classes often only have a few instances with a high degree of visual facial variability. The current techniques of deep representation learning typically implement classic schemes of cost-sensitive or re-sampling learning. Huang et al. [146] (2019) studied the effectiveness of these strategies schemes on class-imbalanced data by employing the learned feature representation. The proposed approach, known as cluster-based large margin local embedding (CLMLE), keeps inter-cluster angular margins between and within classes, thus carving locally more balanced class boundaries.

One of the new ideas in deep face recognition is improving occlusions on variable facial areas, as introduced by Song et al. [147] (2019). The authors have proposed a pairwise differential siamese network (PDSN) framework to find correspondence between corrupted feature elements and occluded facial blocks for deep CNN models resulting in a robust face recognition system under occlusions.

Wei et al. [148] (2020) proposed to solve the problem of margin bias by introducing a minimum margin for full pairs of classes. They presented a loss function called minimum margin loss (MML), which aims to enlarge the overclose class center pairs' margin to enhance the discriminative ability of deep features. MML supervises the training process in conjunction with center loss and Softmax loss to balance all class margins irrespective of their class distributions.

Sun et al. [149] (2020) suggested a novel loss function called inter-class angular margin (IAM) loss, aimed to enlarge inter-class variation adaptively by penalizing smaller inter-class angles more heavily and successfully making the angular margin larger between classes, which can significantly increase the facial features discrimination. The IAM loss is intended to act as a regularization term for the commonly used Softmax loss and its recent variations. Additionally, the authors provided an analysis of the hyper-parameter range of regularization and its effects.

Wu et al. [150] (2020) investigated the impact of quantization errors on face recognition and proposed rotation consistent margin (RCM) loss for efficient low-bit face recognition training by minimizing individual errors, which are necessary to feature discriminative power.

In order to learn the global feature relationships of aligned facial images, Ling et al. [151] (2020) proposed an attention-based neural network (ACNN) for embedding discriminative facial feature, which intends to reduce the information redundancy between channels and concentrate on the most informative components of facial feature maps. The proposed attention module is composed of two blocks called channel attention block and spatial attention block.

To eliminate the large intra-class variance of softmax loss, Wu and Wu [152] (2020) introduced the constraint of cosine similarity into the training process. Two useful loss functions have been proposed named large margin Cosine (LMC) and discriminative large margin Cosine (DLMC). LMC imposes the intraclass cosine similarity between a sample and the corresponding weight vector in the last inner-product layer higher than a given margin. DLMC maintains the inter-class separability and the intra-class compactness simultaneously in the normalized feature space. The proposed loss functions can enhance the deeply learned discriminability. Specifically, as a specialized discriminative large margin Cosine (SDLMC), which has proven to be a variant of triplet loss and presents the intrinsic advantage over the facial verification issue.

Table 3 summarizes all the above works in chronological order evaluated on the LFW dataset, including published time information, network design, number of networks, metric learning, training set, and accuracy.

Table 3. Comparative summary of different deep face verification approaches on the Labeled Face in the Wild (LFW) database.

	Method	Authors	Year	Architecture	Networks	Verif. Metric	Training Set	Accuracy (%) ± SE
1	DeepFace	Taigman et al. [15]	2014	CNN-9	3	Softmax	Facebook (4.4 M, 4 K) *	97.35 ± 0.25
2	DeepID	Sun et al. [116]	2014	CNN-9	60	Softmax + JB	CelebFaces + [116] (202 k, 10 k) *	97.45 ± 0.26
3	DeepID2	Sun et al. [117]	2014	CNN-9	25	Contrastive Softmax + JB	CelebFaces+ (202 k, 10 k) *	99.15 ± 0.13
4	DeepID2+	Sun et al. [118]	2014	CNN-9	25	Contrastive Softmax + JB	WDRef [153] + CelebFaces + (290 k, 12 k) *	99.47 ± 0.12
5	DeepID3	Sun et al. [119]	2015	VGGNet	25	Contrastive Softmax + JB	WDRef + CelebFaces + (290 k,12 k)	99.53 ± 0.10
6	FaceNet	Schroff et al. [127]	2015	GoogleNet	1	Triplet Loss	Google (200 M, 8 M) *	99.63 ± 0.09
7	Web-Scale	Taigman et al. [120]	2015	CNN-9	4	Contrastive Softmax	Private Database (4.5 M, 55 K) *	98.37
8	BAIDU	Liu et al. [122]	2015	CNN-9	10	Triplet Loss	Private Databse (1.2 M, 18 K) *	99.77
9	VGGFace	Parkhi et al. [38]	2015	VGGNet	1	Triplet Loss	VGGFace (2.6 M, 2.6 K)	98.95
10	Augmentation	Masi et al. [123]	2016	VGGNet-19	1	Softmax	CASIA WebFace (494 k, 10 k)	98.06
11	Range Loss	Zhang et al. [124]	2016	VGGNet-16	1	Range Loss	CASIA WebFace + MS-Celeb-1M (5 M, 100 k)	99.52
12	Center Loss	Wen et al. [130]	2016	LeNet	1	Center Loss	CASIA WebFace + CACD2000 [154] + Celebrity + [155] (0.7 M, 17 k)	99.28
13	L-Softmax	Liu et al. [125]	2016	VGGNet-18	1	L-Softmax	CASIA-WebFace (490 k, 10 K)	98.71
14	L2-Softmax	Ranjan et al. [133]	2017	ResNet-101	1	L2-Softmax	MS-Celeb 1M (3.7 M, 58 k)	99.78
15	Marginal Loss	Deng et al. [134]	2017	ResNet-27	1	Marginal Loss	MS-Celeb 1M (4 M, 82 k)	99.48
16	NormFace	Wang et al. [135]	2017	ResNet-28	1	Contrastive Loss	CASIA WebFace (494 k, 10 k)	99.19 ± 0.008
17	Noisy Softmax	Chen et al. [126]	2017	VGGNet	1	Noisy Softmax	CASIA WebFace (400 K, 14 k)	99.18
18	COCO Loss	Liu et al. [136]	2017	ResNet-128	1	COCO Loss	MS-Celeb 1M (3 M, 80 k)	**99.86**
19	Center Invariant Loss	Wu et al. [131]	2017	LeNet	1	Center Invariant Loss	CASIA WebFace (0.45 M, 10 k)	99.12
20	Von Mises-Fisher	Hasnat et al. [137]	2017	ResNet-27	1	vMF Loss	MS-Celeb-1M (4.61 M, 61.24 K)	99.63
21	SphereFace	Liu et al. [138]	2018	ResNet-64	1	A-Softmax	CASIA WebFace (494 k, 10 k)	99.42
22	Ring Loss	Zheng et al. [139]	2018	ResNet-64	1	Ring Loss	MS-Celeb-1M (3.5 M, 31 K)	99.50
23	MLR	Guo and Zhang [140]	2018	ResNet-34	1	CCS Loss	MS-Celeb-1M (10 M, 100 K)	99.71
24	Cosface	Wang et al. [141]	2018	ResNet-64	1	Large Margin Cosine Loss	CASIA WebFace (494 k, 10 k)	99.73
25	AM-Softmax	Wang et al. [142]	2018	ResNet-20	1	AM-Softmax Loss	CASIA WebFace (494 k, 10 k)	99.12
26	Light-CNN	Wu et al. [143]	2018	ResNet-29	1	Softmax	MS-Celeb-1M (5 M, 79 K)	99.33
27	Affinity Loss	Hayat et al. [144]	2019	ResNet-50	1	Affinity Loss	VGGFace2 (3.31 M, 8 K)	99.65
28	ArcFace	Deng et al. [145]	2019	ResNet-100	1	ArcFace	MS-Celeb-1M (5.8 M, 85 k)	99.83
29	CLMLE	Huang et al. [146]	2019	ResNet-64	1	CLMLE Loss	CASIA WebFace (494 k, 10 k)	99.62
30	PDSN	Song et al. [147]	2019	ResNet-50	1	Pairwise Contrastive Loss	CASIA WebFace (494 k, 10 k)	99.20
31	Feature Transfer	Yin et al. [132]	2019	LeNet	1	Softmax	MS-Celeb-1M (4.8 M, 76.5 K)	99.55

Table 3. *Cont.*

	Method	Authors	Year	Architecture	Networks	Verif. Metric	Training Set	Accuracy (%) ± SE
32	Ben Fredj work	Ben Fredj et al. [129]	2020	GoogleNet	1	Softmax with center loss	CASIA WebFace (494 k, 10 k)	99.2 ± 0.04
33	MML	Wei et al. [148]	2020	Inception ResNet-V1 [156]	1	MML Loss	VGGFace2 (3.05 M, 8 K)	99.63
34	IAM	Sun et al. [149]	2020	Inception ResNet-V1	1	IAM loss	CASIA WebFace (494 k, 10 k)	99.12
35	RCM loss	Wu et al. [150]	2020	ResNet-18	1	Rotation Consistent Margin loss	CASIA WebFace (494 k, 10 k)	98.91
36	ACNN	Ling et al. [151]	2020	ResNet-100	1	ArcFace Loss	DeepGlint-MS1M (3.9 M, 86 K)	99.83
37	LMC SDLMC DLMC	Wu and Wu [152]	2020	ResNet32	1	LMC loss SDLMC loss DLMC loss	CASIA WebFace (494 k, 10 k)	98.1399.0399.07

JB: Joint Bayesian. *: Private Database. DeepGlint-MS1M: is a well-cleaned version of CASIA-WebFace [30] and MS-Celeb-1M [36] provided by DeepGlint corporation. The best accuracy is underlined and highlighted in Bold.

Academic community made great efforts in developing multiple methods and adopt different network architectures that significantly enhanced deep face recognition performance accuracy. Several promising ideas have been explored to bring advances in CNNs, such as the use of proper activation [143] and various loss functions [15,124–126,130,134,136,138,141,142,145,148–152], the use of metric learning algorithms [38,127], normalization of features and weights [133,135,139], extraction of appearance variation features [123,129,147], use of multi-CNNs to extract features from various facial regions [116–120,122], and other ideas for the issue of imbalanced training data [131,132,140,144,146]. The famous LFW benchmark results continue to climb as more deep face methods are introduced; for example, in pasting four years, the accuracy has been increased from 97.35% with DeepFace (2014) to 99.86% with COCO loss (2017), as mentioned in Table 3. We can deduce that the accuracy of LFW has got saturated, and all the rivals can reach a high accuracy rate.

In summary, deep convolutional neural networks have provided tremendous face recognition by learning more discriminative features on extensive datasets and outperformed recognition performances compared to holistic, geometric, and local-texture approaches. They also showed robustness to variations in pose, orientation, partial occlusion, misalignment, and expression. Although significant improvements have been made with deep learning-based face recognition, there are some challenges: the efficient training of CNN requires large-scale training data, demands hardware advancements such as GPUs, and needs lots of high-quality data.

6. Three-Dimensional Face Recognition

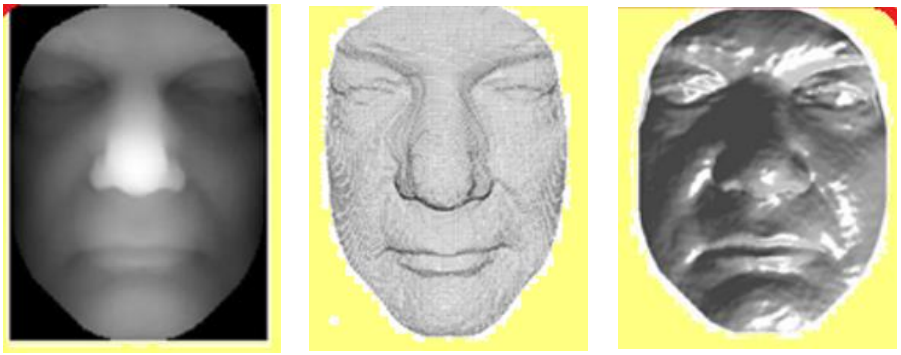
6.1. Factual Background and Acquisition Systems

6.1.1. Introduction to 3D Face Recognition

2D facial recognition systems are limited by constraints such as changes in physical appearance, aging factor, pose, changes in light intensity, and more generally by facial expressions, missing data, cosmetics and occlusions. To overcome these difficulties, 3D facial recognition systems have been developed with the aim of theoretically providing a high level of precision and reliability, and greater immunity to variations in the face due to different factors. Such a capacity is due to more elaborate acquisition systems and to 3D models taking into account the geometric information [157,158].

Face recognition acquisition devices can be a 2D, 3D, or infrared camera or a combination of these modalities. Pre-processing can detect facial landmarks, align facial data, and crop the facial area. It can filter irrelevant information such as hair, background, and reduce facial variations due to the change in pose. In 2D images, landmarks such as eyes, eyebrows, mouth, etc., can be reliably detected, while the nose is the most important landmark in 3D facial recognition. The 3D information (depth and texture maps) corresponding to the surface of the face can be acquired using different alternatives: a multi-camera system (stereoscopy), remote cameras or laser devices, and 3D scanner.

The formation of 3D facial images requires particular hardware devices, which can be classified according to the strategies employed into active and passive acquisition devices [159,160]. The first type is based on the emission/reflection of a non-visible light to illuminate the object and to capture its shape features. According to the various forms of lighting techniques, the active acquisition devices can be moreover grouped into triangulation and structured light techniques. On the other hand, the construction of the 3D facial images with passive acquisition devices is based on the placement of several cameras at predefined places and matching a set of canonical points observed from the installed cameras. Figure 28 shows examples of 3D facial images acquired with triangulation, structured light, and passive acquisitions devices, respectively [159,160].



(**a**) (**b**) (**c**)

Figure 28. Three widely used 3D scanners. (**a**). Example of a 3D facial image acquired with a triangulation-based device, known as depth image, (**b**). example of a 3D facial image acquired with a structured light-based device, known as point cloud, (**c**). example of a 3D facial image acquired with passive acquisition device, known as mesh.

6.1.2. Microsoft Kinect Technology

Among active acquisition systems based on structured light technology, emerging RGB-D (red green blue-depth) cameras such as the Microsoft Kinect sensor are beginning to be successfully applied to 3D facial recognition [161]. The choice of Microsoft Kinect is motivated by its efficiency, its low cost, its ease of RGB-D mapping, and multimodal detection.

The original version of Microsoft Kinect sensor consists of a RGB camera, an infrared camera, an IR projector, a multi-array microphone, and a motorized tilt (see Figure 29). Figure 30 shows the acquisition environment for the Kinect face database. Figure 31 shows two example images captured by depth sensors and RGB camera, respectively. Here, RGB camera is able to provide the image with the resolution of 640 × 480 pixels at 30 Hz. This RGB camera also has option to produce higher resolution images (1280 × 1024 pixels), running at 10 Hz. Kinect's 3D depth sensor (infrared camera and IR projector) can provide depth images with the resolution of 640 × 480 pixels at 30 Hz.

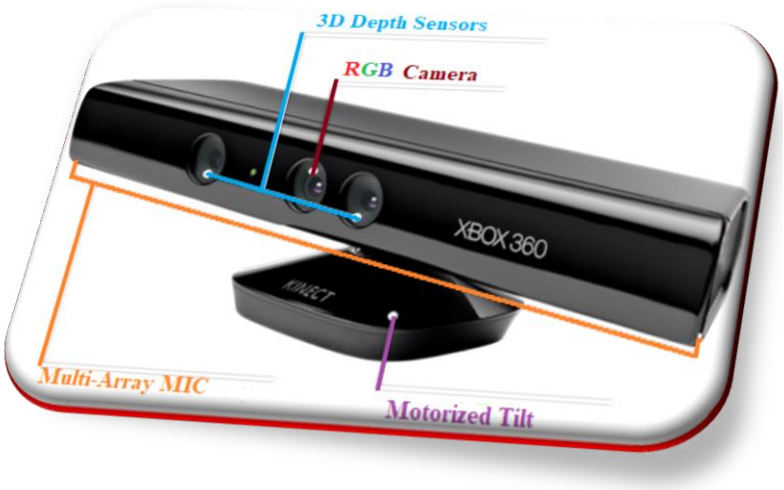
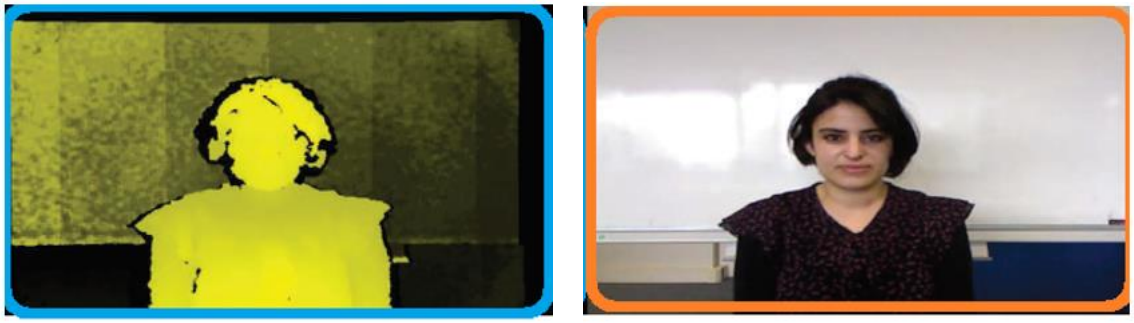


Figure 29. Microsoft Kinect Sensor.

Figure 30. Acquisition environment for the Kinect face database.



(**a**) (**b**)

Figure 31. RGB-D (red green blue-depth) alignment: (**a**) the depth map is aligned with (**b**) the RGB image captured by Kinect at the same time.

The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions.

Kinect technology has just been introduced toward the end of 2017 in a revolutionary smartphone: iphone X.

The iPhone X (pronounced iPhone 10 for the Roman numeral X which represents the tenth anniversary of the iPhone1) is a model of the 11th generation of the smartphone from the company Apple (see Figure 32). It marks a break with the older generations of iPhone mostly with its design incorporating a "borderless" super retina screen (without border and without home button) with the highest resolution ever seen on an iPhone and also thanks to its 3D capture technology with the "TrueDepth" camera which allows in particular the integration of an "invisible" secure unlocking technology: Face ID. The iPhone X projects 30,000 infrared points to create an embossed mold on the user's face. This technology is the most advanced means of security on a smartphone, with a failure rate (1/1,000,000). It is therefore the first smartphone to have 3D facial recognition technology. The classic method of facial recognition (used on other smartphones) uses the front camera, security that can be deceived with a simple picture.

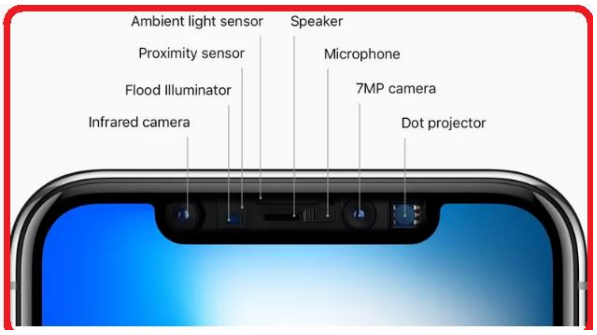


Figure 32. The iPhone X's notch is basically a Kinect.

6.2. Methods and Datasets

6.2.1. Challenges of 3D Facial Recognition

3D face recognition takes advantage of the 3D geometric details of the human face. It uses 3D sensor data to collect details on the shape of a face, and recognition is based on matching metadata of the 3D shape of the face.

The 3D capture process is becoming cheaper, more accessible, and faster, which is why many contributions have been made in the last ten years to improve facial recognition based on a 3D facial model. The community of researchers in this field has intensively explored 3D face recognition in order to solve three main unsolved problems in 2D face recognition such as sensitivity to light conditions (Figure 33), pose (Figure 34), and use of makeup or beautification (Figure 35). Although 3D facial representations are theoretically insensitive to lighting variations, they must still be processed correctly before the matching process.

Figure 33. Lighting variation. The original image is on the left.

Figure 34. Change of pose. The original image is on the left.

(**a**) (**b**) (**c**)

Figure 35. (**a**,**b**) Plastic surgery. (**c**) Facial cosmetics.

In real life, it is very likely that certain parts of the face are obstructed by sunglasses, a hat, a scarf, mask, hands moving over the mouth, a mustache or hair, etc. This is called partial occlusion presented in the face or simply occlusion. Figure 36 shows an example of occlusion from the Kinect face database [161]. Occlusion can significantly alter the visual appearance of the face and therefore seriously degrade the performance of facial recognition systems. In addition, the problem of influences of facial expressions such as anger, disgust, fear, happiness, sadness, and surprise (Figure 37), and more generally emotions (Figure 38) is an open challenge because of the complexity of the 3D model.

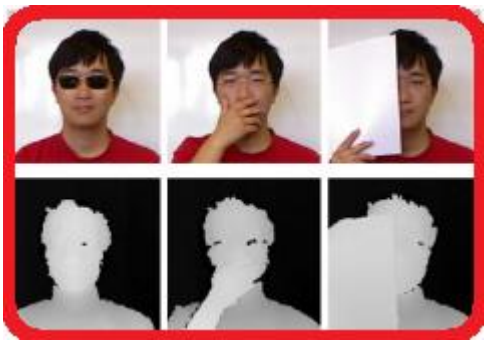


Figure 36. Occlusions by sunglasses, by hand and by paper. Upper: the RGB (red green blue) images. Lower: the depth maps aligned with above RGB images.

Figure 37. Facial expressions.

Occlusion and pose variation issues have been the subject of work since using the GavabDB, Bosphorus and FRGCv2 databases (see Table 4) where the estimation of missing facial parts uses PCA on tangent spaces and by calculating mean shapes [162].

For the problems of 3D facial expression recognition, we recommend to refer to Alexandre's systematic review [163] which reveals that after pre-processing and machine learning the expressions of happiness and surprise are the most regularly distinguished, while fear and sadness turned out to be the most difficult expressions, thus representing an opportunity for future dedicated work.

Figure 38. Facial emotion images.

Table 4. Comparative summary of some accessible 3D face recognition databases.

Database	Apparition's Date	Images	Subjects	Data Type
BU-3DFE	2006	2500	100	Mesh
FRGC v1.0 [14]	2006	943	273	Depth image
FRGC v2.0 [14]	2006	4007	466	Depth image
CASIA	2006	4623	123	Depth image
ND2006	2007	888	13,450	Depth image
Bosphorus	2008	4666	105	Point Cloud
BJUT-3D	2009	1200	500	Mesh
Texas 3DFRD	2010	1140	118	Depth image
UMB-DB	2011	1473	143	Depth image
BU-4DFE	2008	606 sequences = 60,600 (frames)	101	3D video

6.2.2. Traditional Methods of Machine Learning

Three-dimensional facial recognition can be performed using one of the following two strategies:

- Traditional methods of machine learning
- Deep learning-based methods.

Traditional methods are generally divided into three categories: holistic, local, and hybrid approaches. In the holistic approach, the focus is on the similarity of faces. The entire 3D face is described by defining a set of global features. Principal component analysis and deformation modeling are the most popular holistic methods. The local approach examines the geometric features of the face,

mainly the eyes and the nose. The hybrid solution integrates holistic as well as local characteristics or data (2D and 3D images).

Although several studies have been carried out using holistic methods, it seems that local methods are more suitable for recognizing faces in 3D. Compared to holistic methods, local methods are more robust in terms of occlusion and can give better experimental results [159]. However, if the face is frontal, and there is no variation in expression, the hybrid solution is very effective.

Figure 39 illustrates the 3-dimensional facial recognition process using traditional machine learning methods.

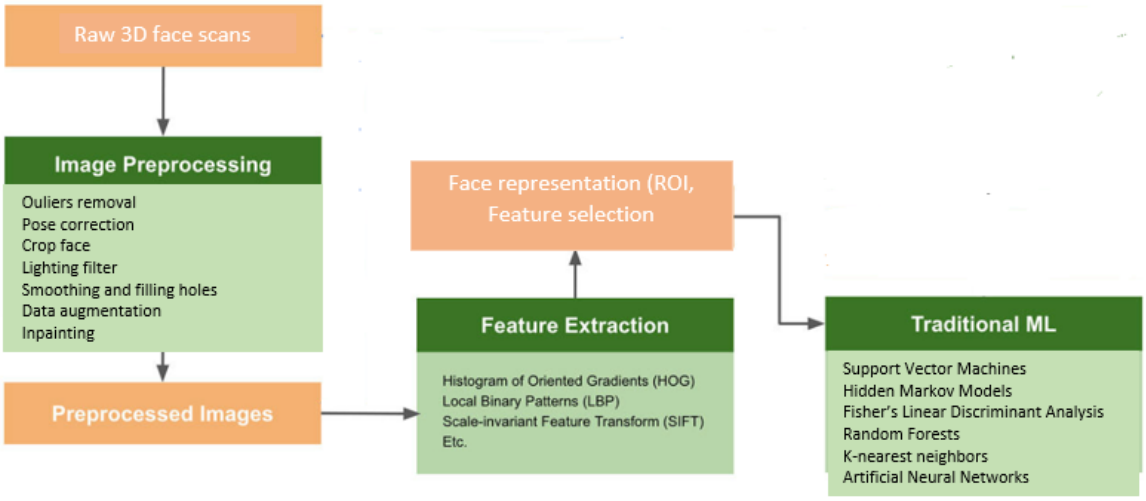


Figure 39. Summary of the usual pipeline in 3D face recognition.

6.2.3. Deep Learning-Based Methods

The various methods based on deep learning (see Section 5.4) in 3D facial recognition represent less than 10% of work in this area over the past 5 years.

In theory, these methods are efficient and do not require the definition of a region of interest (ROI), nor the extraction and selection of features.

However, some artisanal facial representation operations can precede the deep learning network. This makes real-time and immediate 3D facial recognition somewhat laborious. In practice, deep learning requires a learning process on a large volume of data, this is the concept of big data. However, the number of 3D facial scans available is very limited, which makes recognition performance in terms of accuracy (recognition rate) very critical and unreliable.

To clarify this point, for example, the FaceNet dataset [164] used in 2D facial recognition by deep learning is very large of the order of 200 M for training a deep CNN, while in 3D facial recognition, the best datasets contain 2 k to 15 k. For example, the famous Bosphorus dataset contains around 4 k images and BU-3DFE contains around 2.5 k (see Table 4). It is therefore clear that the performance of 3D facial recognition is below the expectations of its promoters even if some authors [165–168] obtain relatively acceptable performance but at the cost of complications and rather confused preprocessing.

6.2.4. Three-Dimensional Face Recognition Databases

2D facial recognition methods based on deep CNN extractors trained on a massive dataset outperform conventional methods using classical feature extractors, such as support vector machines, hidden Markov models, random forests, K-nearest neighbors, Fisher's linear discriminant analysis, artificial neural networks.

Although 3D facial recognition based on deep learning is very difficult because of the lack of large scale 3D facial datasets, 3D models have the potential address changes in texture, expression, pose, and face scaling, which is not the case with 2D data.

Besides, some problems are not yet solved well, mainly when the subject is non-cooperative during the acquisition process, which can cause a difference in the posture, facial expression, and generates occlusions, by foreign bodies on the facial surface [159,160].

Also, interpretation of the 3D facial expression, identification under variations in age, and transfer learning are three open challenges that are still in their beginning and requires further researches.

In the database side, we can note that the academic community has plenty of large-scale 2D facial databases. These databases provide an official forum for assessing and comparing face recognition algorithms in 2D. However, databases with 3D faces are less frequent and smaller in size because building 2D facial data can be obtained easily by searching the Internet, while 3D facial data involves physical collection from real subjects, restricting its use and evolution. Some of the well-known available 3D face recognition datasets are described in Table 4, which compares various types of data formats, the number of images, the number of subjects, and the date of the apparition.

The BU-3DFE and Bosphorus databases are currently the stars of 3D facial recognition studies.

As summarized in Table 4, BU-3DFE database consists of 2500 scans acquired from 100 subjects where each subject offers a neutral pose and six basic emotions with 4 nuances or degrees (Figures 37 and 38 show several degrees of happiness, anger...).

Bosphorus is a 3D facial recognition database (and even 3D facial expression recognition) widely used in validation. Table 4 shows that Bosphorus contains 105 subjects, of which only 65 have facial expressions.

It is clear that the databases mentioned in Table 3 are not consistent in terms of the number of scans and pose problems when methods based on deep learning are tried to be applied.

The importance of creating face recognition datasets is essential, first, for security-related applications, and second, to allow the development and validation of methods based on deep learning in 3D facial recognition. Thus, even in specific fields such as autonomous vehicles, a multimodal database has recently been proposed [169] and can be supplemented by 3D facial expression recognition.

7. Open Challenges

As in most biometric applications, appearance variations caused by unconstrained environments tend to present open face recognition challenges. In the following paragraphs, we will cite some challenges to be met in the near future.

7.1. Face Recognition and Occlusion

The face can be captured in an arbitrary pose in a specific environment and without any user support, so it is possible that the image contains only a partial face. Facial occlusions can be, for example, shades, scarf, hat and veil, facial artifacts (for example, hand, food and cell phone), bright light (for example, shadow), self-occlusion (for example, non-frontal pose) or poor image quality (for example, blurring), as shown in Figure 40. For example, in forensic face identification, a suspect must be identified in the crowd by matching a partially occluded face based on a recorded image. There is a two-fold difficulty in recognizing facial occlusion. First, the occlusion distorts the discriminating facial features and increases the distance in the feature space between two images of the same object. The intra-class variations are larger than the inter-class variations, contributing to low results in recognition. Second, significant alignment errors typically occur when facial landmarks are occluded and degrade recognition rates [170].

Figure 40. Some examples of occlusion by hat, glass, mask, hand, shadow, and self-occlusion.

7.2. Hetegerenous Face Recognition

Recognition of heterogeneous faces involves the correlation between two facial representations from different imaging methods; this is very useful in legal situations. For example, infrared imaging [171,172] (Figure 41 on the right) may be the only way to acquire a suspect's useful image in night environments, but the police recorded files are visible images. A further example is a correspondence with sketch photographs (in the center of Figure 41); when no suspect image is available, a legal sketch is created based on an eyewitness description. Correspondence from sketches against facial photographs is essential in legal inquiries.

(**a**) (**b**) (**c**)

Figure 41. Some modalities of imaging display heterogeneous faces. (**a**) Simple photographs. (**b**) Sketch images. (**c**) Infrared images.

7.3. Face Recognition and Ageing

Facial aging is a complex process that affects a face's form and texture (e.g., skin tone or wrinkles). The typical scenario for applying face recognition systems against the effect of aging is to detect the presence of a particular person in a previously registered database (e.g., the identification of missing children or control of suspects on a watch list). As the age between a query image and a reference image of the same person increases, it generally decreases the accuracy of recognition systems (Figure 42) [173].

Figure 42. Example of facial aging.

7.4. Single Sample Face Recognition

One of the most exclusive and realistic situations of applying face recognition is the single sample per person (SSPP), or merely single sample face recognition (SSFR) [9] (Figure 43). It is one of the most challenging facial recognition issues, where there is only one facial representation per individual for training-set. It is well-known that:

- In real-world applications (e.g., passports, immigration systems), only one model of each individual is registered in the database and accessible for the recognition task [174].

- Pattern recognition systems require vast training data to ensure the generalization of the learning systems.
- Deep learning-based approach is considered a powerful technique in face recognition. Nonetheless, they need a significant amount of training data to perform well [9].

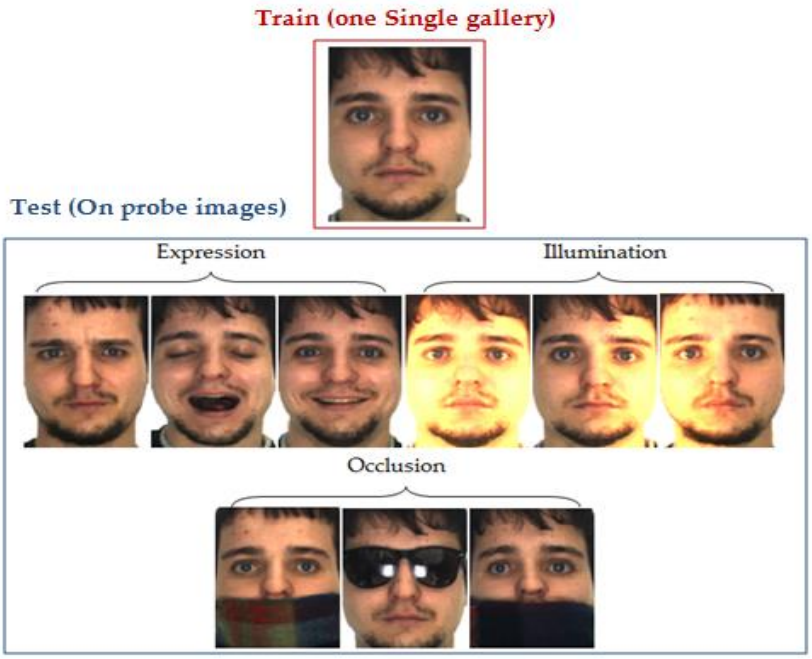


Figure 43. Example of face recognition by single sample per person.

In summary, we may conclude that SSFR remains an unresolved issue and is among the most common subjects in academia or industry.

7.5. Face Recognition in Video Surveillance

Face-based recognition systems are becoming frequently common and find very varied applications, especially in video surveillance [175]. In this setting, facial recognition systems' performance is mostly reliant on image acquisition conditions, mainly when the posture changes, and because the acquisition techniques themselves may include artifacts. So, we are mainly talking about camera focus problems that can lead to image blurring, low-resolution, or compression-related errors and block effects. The challenge of face recognition systems, in this case, is to distinguish individuals from photographs captured employing video surveillance cameras, presenting blurred, low-resolution, block artifacts, or faces with variable poses (Figure 44). This challenge remains an unsolved problem and requires further research.

7.6. Face Recognition and Soft Biometrics

Soft biometrics is a human characteristic providing specific information that can be used to determine, for example, age, gender, eye and hair color, height, weight, skin color, or ethnicity of the person from facial images. Unlike hard biometrics, which consists of determining the person's identity based on distinct and permanent personal features from facial images [16], soft biometrics offers ambiguous information that is not necessarily permanent or distinguishable. These soft features are usually more natural to capture remotely and do not require object co-operation. While they cannot provide reliable authentication, they can be used as additional information to reduce matching operations, which will improve the recognition performance of hard face recognition systems [176].

Despite the vast potential applications that have been made in this context, soft biometrics research is still in its infancy and needs further research.

Figure 44. Examples of face recognition in video surveillance.

Figure 45 shows an example of soft biometrics: facial marks (e.g., freckles, scars, moles, tattoos, chipped teeth, lip creases . . .) used to improve face recognition.

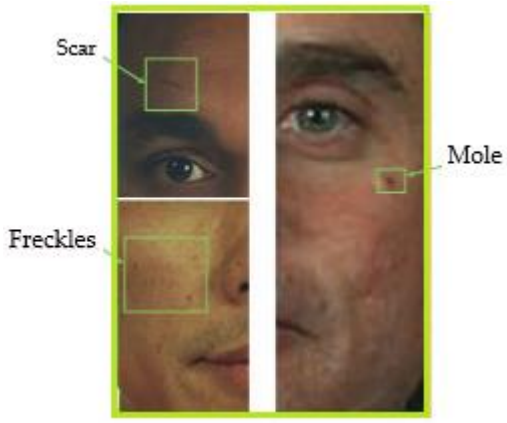


Figure 45. Soft biometrics: facial marks (freckles, mole, and scar).

Although these micro-features cannot uniquely identify an individual, they can restrict the search for an identity.

7.7. Face Recognition and Smartphones

Adopting face recognition on mobile devices offers many advantages. In addition to the employment facility, the users do not have to remember either the PIN or password; it can be conveniently implemented on tablets and smartphones because only the frontal camera is required. Face recognition systems have been used in recent years to secure the devices and control access to many different services through smartphones, such as a purchase or online payments on the store, as shown in Figure 46.

Figure 46. Examples of face recognition applications using smartphones.

While the adoption of face recognition systems on smartphones provides many advantages, many challenges need to be addressed. The facial image of the user should be captured in a comfortable or constrained environment. Many factors, such as pose and ambient lighting due to various ways of interacting with mobile technology and imaging distance, can restrict facial image quality [177].

7.8. Face Recognition and Internet of Things (IoT)

While there are several risks with facial recognition, it also offers numerous solutions for future and upcoming technologies. Currently the internet of things (IoT) technology is booming as well as the way to connect domestic or urban devices to the Internet to make them "smart" [178].

By integrating facial recognition into the IoT, various simplifications of life will be available. For example, the door of an apartment could recognize the resident and open automatically. A more common example is the simple activation of the smartphone by facial recognition via its front camera (Figure 47a). Similarly, this technology can help us to find lost relatives when visiting an unknown city (Figure 47b). A program will search for correspondence to the person's profile photo in a predefined database by uploading a photo to a website. Finally, facial recognition can increase security by identifying criminals, assuming that this technology's accuracy rate is nearly 100%.

(**a**) (**b**)

Figure 47. Examples of IoT (internet of things) using face recognition: (**a**) door unlock and (**b**) smart city.

8. Conclusions

This systematic review provides the new state-of-the-art in facial recognition research in a comprehensive manner. Recent advances in this field are clearly stated and prospects for improvement are proposed.

The outcomes of this review show that a substantial boost in this domain's research occurred over the last five years, particularly with the advent of deep learning approach that has outperformed the most popular computer vision methods. In addition, numerous facial databases (public and private) are available for research and commercial purposes and their main characteristics and evaluation protocols are presented. A focus on the labeled faces in the wild (LFW) database in terms of methodology, architecture, metrics, precision, and protocols was necessary to allow researchers to compare their results to this referential database.

The main lessons learnt from this study are that 2D facial recognition is still open to future technical and material developments for the acquisition of images to be analyzed. On the other hand, the attention of researchers is increasingly attracted by 3D facial recognition. The recent development of 3D sensors reveals a new direction for facial recognition that could overcome the main limitations of 2D technologies, e.g., changes in physical appearance, aging factor, pose, changes in light intensity, and more generally by facial expressions, missing data, cosmetics, and occlusions. The geometric information provided by 3D facial data could significantly improve the accuracy of facial recognition in the presence of adverse acquisition conditions. However, the lack of a 3D facial recognition database hinders the exploitation of methods based on deep learning. Also, interpretation of the 3D facial expression, identification under variations in age, and transfer learning are three open challenges that are still in their beginning and requires further researches.

Multimodality (voice, iris, fingerprint, ...), soft facial biometrics, infrared imaging, sketches, and deep learning without neglecting conventional machine learning methods are tracks to be considered in the near future.

Naturally, these new developments in facial recognition must meet four objectives: always faster (immediate response seen from the user's point of view), accuracy close to 100%, optimal security, miniaturized, and portable equipment.

Author Contributions: Software and writing, I.A.; methodology, validation, and review, A.O.; investigation and writing, A.B.; supervision and review, A.T.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. A Review of Face Recognition Methods. *Sensors* **2020**, *20*, 342. [CrossRef] [PubMed]
2. O'Toole, A.J.; Roark, D.A.; Abdi, H. Recognizing moving faces: A psychological and neural synthesis. *Trends Cogn. Sci.* **2002**, *6*, 261–266. [CrossRef]
3. Dantcheva, A.; Chen, C.; Ross, A. Can facial cosmetics affect the matching accuracy of face recognition systems? In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; pp. 391–398.
4. Sinha, P.; Balas, B.; Ostrovsky, Y.; Russell, R. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* **2006**, *94*, 1948–1962. [CrossRef]
5. Ouamane, A.; Benakcha, A.; Belahcene, M.; Taleb-Ahmed, A. Multimodal depth and intensity face verification approach using LBP, SLF, BSIF, and LPQ local features fusion. *Pattern Recognit. Image Anal.* **2015**, *25*, 603–620. [CrossRef]
6. Porter, G.; Doran, G. An anatomical and photographic technique for forensic facial identification. *Forensic Sci. Int.* **2000**, *114*, 97–105. [CrossRef]
7. Li, S.Z.; Jain, A.K. *Handbook of Face Recognition*, 2nd ed.; Springer Publishing Company: New York, NY, USA, 2011.
8. Morder-Intelligence. Available online: https://www.mordorintelligence.com/industry-reports/facial-recognition-market (accessed on 21 July 2020).
9. Guo, G.; Zhang, N. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.* **2019**, *189*, 10285. [CrossRef]
10. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; Technical Report; University of Massachusetts: Amherst, MA, USA, 2007; pp. 7–49.
11. Bledsoe, W.W. *The Model Method in Facial Recognition*; Technical Report; Panoramic Research, Inc.: Palo Alto, CA, USA, 1964.
12. Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86. [CrossRef]

13. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P. The FERET database and evaluation procedure for face recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [CrossRef]
14. Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W. Overview of the face recognition grand challenge. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; pp. 947–954.
15. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
16. Chihaoui, M.; Elkefi, A.; Bellil, W.; Ben Amar, C. A Survey of 2D Face Recognition Techniques. *Computers* **2016**, *5*, 21. [CrossRef]
17. Benzaoui, A.; Bourouba, H.; Boukrouche, A. System for automatic faces detection. In Proceedings of the 2012 3rd International Conference on Image Processing, Theory, Tools and Applications (IPTA), Istanbul, Turkey, 15–18 October 2012; pp. 354–358.
18. Martinez, A.M. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **2002**, *24*, 748–763. [CrossRef]
19. Sidahmed, S.; Messali, Z.; Ouahabi, A.; Trépout, S.; Messaoudi, C.; Marco, S. Nonparametric denoising methods based on contourlet transform with sharp frequency localization: Application to electron microscopy images with low exposure time. *Entropy* **2015**, *17*, 2781–2799.
20. Ouahabi, A. Image Denoising using Wavelets: Application in Medical Imaging. In *Advances in Heuristic Signal Processing and Applications*; Chatterjee, A., Nobahari, H., Siarry, P., Eds.; Springer: Basel, Switzerland, 2013; pp. 287–313.
21. Ouahabi, A. A review of wavelet denoising in medical imaging. In Proceedings of the International Workshop on Systems, Signal Processing and Their Applications (IEEE/WOSSPA'13), Algiers, Algeria, 12–15 May 2013; pp. 19–26.
22. Nakanishi, A.Y.J.; Western, B.J. Advancing the State-of-the-Art in Transportation Security Identification and Verification Technologies: Biometric and Multibiometric Systems. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, 30 September–3 October 2007; pp. 1004–1009.
23. Samaria, F.S.; Harter, A.C. Parameterization of a Stochastic Model for Human Face Identification. In Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 138–142.
24. Martinez, A.M.; Benavente, R. The AR face database. *CVC Tech. Rep.* **1998**, *24*, 1–10.
25. Messer, K.; Matas, J.; Kittler, J.; Jonsson, K. Xm2vt sdb: The extended m2vts database. In Proceedings of the 1999 2nd International Conference on Audio and Video-based Biometric Person Authentication (AVBPA), Washington, DC, USA, 22–24 March 1999; pp. 72–77.
26. Bailliére, E.A.; Bengio, S.; Bimbot, F.; Hamouz, M.; Kittler, J.; Mariéthoz, J.; Matas, J.; Messer, K.; Popovici, V.; Porée, F.; et al. The BANCA Database and Evaluation Protocol. In Proceedings of the 2003 International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), Guildford, UK, 9–11 June 2003; pp. 625–638.
27. Huang, G.B.; Jain, V.; Miller, E.L. Unsupervised joint alignment of complex images. In Proceedings of the 2007 IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 14–20 October 2007; pp. 1–8.
28. Huang, G.; Mattar, M.; Lee, H.; Miller, E.G.L. Learning to align from scratch. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 764–772.
29. Gross, R.; Matthews, L.; Cohn, J.; Kanade, T.; Baker, S. Multi-PIE. *Image Vis. Comput.* **2010**, *28*, 807–813. [CrossRef] [PubMed]
30. CASIA Web Face. Available online: http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html (accessed on 21 July 2019).
31. Klare, B.F.; Klein, B.; Taborsky, E.; Blanton, A.; Cheney, J.; Allen, K.; Grother, P.; Mah, A.; Burge, M.; Jain, A.K. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1931–1939.

32. Shlizerman, I.K.; Seitz, S.M.; Miller, D.; Brossard, E. The MegaFace benchmark: 1 million faces for recognition at scale. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4873–4882.
33. Shlizerman, I.K.; Suwajanakorn, S.; Seitz, S.M. Illumination-aware age progression. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 3334–3341.
34. Ng, H.W.; Winkler, S. A data-driven approach to cleaning large face datasets. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 343–347.
35. Sengupta, S.; Cheng, J.; Castillo, C.; Patel, V.M.; Chellappa, R.; Jacobs, D.W. Frontal to Profile Face Verification in the Wild. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9.
36. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. Ms-Celeb-1m: A dataset and benchmark for large-scale face recognition. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
37. Wang, T.Y.; Kumar, A. Recognizing Human Faces under Disguise and Makeup. In Proceedings of the 2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), Sendai, Japan, 29 February–2 March 2016; pp. 1–7.
38. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the 2015 British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.
39. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A dataset for recognizing faces across pose and age. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG), Xi'an, China, 15–19 May 2018; pp. 67–74.
40. Whitelam, C.; Taborsky, E.; Blanton, A.; Maze, B.; Adams, J.; Miller, T.; Kalka, N.; Jain, A.K.; Duncan, J.A.; Allen, K. IARPA Janus Benchmark-B face dataset. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 592–600.
41. Nech, A.; Shlizerman, I.K. Level playing field for million scale face recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3406–3415.
42. Kushwaha, V.; Singh, M.; Singh, R.; Vatsa, M. Disguised Faces in the Wild. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1–18.
43. Maze, B.; Adams, J.; Duncan, J.A.; Kalka, N.; Miller, T.; Otto, C.; Jain, A.K.; Niggel, W.T.; Anderson, J.; Cheney, J.; et al. IARPA Janus benchmark-C: Face dataset and protocol. In Proceedings of the 2018 International Conference on Biometrics (ICB), Gold Coast, QLD, Australia, 20–23 February 2018; pp. 158–165.
44. Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S. LFR face dataset: Left-Front-Right dataset for pose-invariant face recognition in the wild. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2–5 February 2020; pp. 124–130.
45. Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. Masked Face Recognition Dataset and Application. *arXiv* **2020**, arXiv:2003.09093v2.
46. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **1997**, *19*, 711–720. [CrossRef]
47. Stone, J.V. Independent component analysis: An introduction. *Trends Cogn. Sci.* **2002**, *6*, 59–64. [CrossRef]
48. Sirovich, L.; Kirby, M. Low-Dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am.* **1987**, *4*, 519–524. [CrossRef]
49. Kirby, M.; Sirovich, L. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **1990**, *12*, 831–835. [CrossRef]
50. Femmam, S.; M'Sirdi, N.K.; Ouahabi, A. Perception and characterization of materials using signal processing techniques. *IEEE Trans. Instrum. Meas.* **2001**, *50*, 1203–1211. [CrossRef]
51. Zhao, L.; Yang, Y.H. Theoretical analysis of illumination in PCA-based vision systems. *Pattern Recognit.* **1999**, *32*, 547–564. [CrossRef]
52. Pentland, A.; Moghaddam, B.; Starner, T. View-Based and modular eigenspaces for face recognition. In Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 21–23 June 1994; pp. 84–91.

53. Bartlett, M.; Movellan, J.; Sejnowski, T. Face Recognition by Independent Component Analysis. *IEEE Trans. Neural Netw.* **2002**, *13*, 1450–1464. [CrossRef]
54. Abhishree, T.M.; Latha, J.; Manikantan, K.; Ramachandran, S. Face recognition using Gabor Filter based feature extraction with anisotropic diffusion as a pre-processing technique. *Procedia Comput. Sci.* **2015**, *45*, 312–321. [CrossRef]
55. Zehani, S.; Ouahabi, A.; Oussalah, M.; Mimi, M.; Taleb-Ahmed, A. Trabecular bone microarchitecture characterization based on fractal model in spatial frequency domain imaging. *Int. J. Imaging Syst. Technol.*. accepted.
56. Ouahabi, A. *Signal and Image Multiresolution Analysis*, 1st ed.; ISTE-Wiley: London, UK, 2012.
57. Guetbi, C.; Kouame, D.; Ouahabi, A.; Chemla, J.P. Methods based on wavelets for time delay estimation of ultrasound signals. In Proceedings of the 1998 IEEE International Conference on Electronics, Circuits and Systems, Lisbon, Portugal, 7–10 September 1998; pp. 113–116.
58. Ferroukhi, M.; Ouahabi, A.; Attari, M.; Habchi, Y.; Taleb-Ahmed, A. Medical video coding based on 2nd-generation wavelets: Performance evaluation. *Electronics* **2019**, *8*, 88. [CrossRef]
59. Wang, M.; Jiang, H.; Li, Y. Face recognition based on DWT/DCT and SVM. In Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM), Taiyuan, China, 22–24 October 2010; pp. 507–510.
60. Bookstein, F.L. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **1989**, *11*, 567–585. [CrossRef]
61. Shih, F.Y.; Chuang, C. Automatic extraction of head and face boundaries and facial features. *Inf. Sci.* **2004**, *158*, 117–130. [CrossRef]
62. Zobel, M.; Gebhard, A.; Paulus, D.; Denzler, J.; Niemann, H. Robust facial feature localization by coupled features. In Proceedings of the 2000 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG), Grenoble, France, 26–30 March 2000; pp. 2–7.
63. Wiskott, L.; Fellous, J.M.; Malsburg, C.V.D. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **1997**, *19*, 775–779. [CrossRef]
64. Xue, Z.; Li, S.Z.; Teoh, E.K. Bayesian shape model for facial feature extraction and recognition. *Pattern Recognit.* **2003**, *36*, 2819–2833. [CrossRef]
65. Tistarelli, M. Active/space-variant object recognition. *Image Vis. Comput.* **1995**, *13*, 215–226. [CrossRef]
66. Lades, M.; Vorbuggen, J.C.; Buhmann, J.; Lange, J.; Malsburg, C.V.D.; Wurtz, R.P.; Konen, W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **1993**, *42*, 300–311. [CrossRef]
67. Wiskott, L. Phantom faces for face analysis. *Pattern Recognit.* **1997**, *30*, 837–846. [CrossRef]
68. Duc, B.; Fischer, S.; Bigun, J. Face authentication with Gabor information on deformable graphs. *IEEE Trans. Image Process.* **1999**, *8*, 504–516. [CrossRef] [PubMed]
69. Kotropoulos, C.; Tefas, A.; Pitas, I. Frontal face authentication using morphological elastic graph matching. *IEEE Trans. Image Process.* **2000**, *9*, 555–560. [CrossRef] [PubMed]
70. Jackway, P.T.; Deriche, M. Scale-space properties of the multiscale morphological dilation-erosion. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **1996**, *18*, 38–51. [CrossRef]
71. Tefas, A.; Kotropoulos, C.; Pitas, I. Face verification using elastic graph matching based on morphological signal decomposition. *Signal Process.* **2002**, *82*, 833–851. [CrossRef]
72. Kumar, D.; Garaina, J.; Kisku, D.R.; Sing, J.K.; Gupta, P. Unconstrained and Constrained Face Recognition Using Dense Local Descriptor with Ensemble Framework. *Neurocomputing* **2020**. [CrossRef]
73. Zehani, S.; Ouahabi, A.; Mimi, M.; Taleb-Ahmed, A. Staistical features extraction in wavelet domain for texture classification. In Proceedings of the 2019 6th International Conference on Image and Signal Processing and their Applications (IEEE/ISPA), Mostaganem, Algeria, 24–25 November 2019; pp. 1–5.
74. Ait Aouit, D.; Ouahabi, A. Nonlinear Fracture Signal Analysis Using Multifractal Approach Combined with Wavelet. *Fractals Complex Geom. Patterns Scaling Nat. Soc.* **2011**, *19*, 175–183. [CrossRef]
75. Girault, J.M.; Kouame, D.; Ouahabi, A. Analytical formulation of the fractal dimension of filtered stochastic signal. *Signal Process.* **2010**, *90*, 2690–2697. [CrossRef]
76. Djeddi, M.; Ouahabi, A.; Batatia, H.; Basarab, A.; Kouamé, D. Discrete wavelet transform for multifractal texture classification: Application to ultrasound imaging. In Proceedings of the IEEE International Conference on Image Processing (IEEE ICIP2010), Hong Kong, China, 26–29 September 2010; pp. 637–640.

77. Ouahabi, A. Multifractal analysis for texture characterization: A new approach based on DWT. In Proceedings of the 10th International Conference on Information Science, Signal Processing and Their Applications (IEEE/ISSPA), Kuala Lumpur, Malaysia, 10–13 May 2010; pp. 698–703.
78. Davies, E.R. Introduction to texture analysis. In *Handbook of Texture Analysis*; Mirmehdi, M., Xie, X., Suri, J., Eds.; Imperial College Press: London, UK, 2008; pp. 1–31.
79. Benzaoui, A.; Hadid, A.; Boukrouche, A. Ear biometric recognition using local texture descriptors. *J. Electron. Imaging* **2014**, *23*, 053008. [CrossRef]
80. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face recognition with local binary patterns. In Proceedings of the 8th European Conference on Computer Vision (ECCV), Prague, Czech Republic, 11–14 May 2004; pp. 469–481.
81. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2037–2041. [CrossRef]
82. Beveridge, J.R.; Bolme, D.; Draper, B.A.; Teixeira, M. The CSU face identification evaluation system: Its purpose, features, and structure. *Mach. Vis. Appl.* **2005**, *16*, 128–138. [CrossRef]
83. Moghaddam, B.; Nastar, C.; Pentland, A. A bayesian similarity measure for direct image matching. In Proceedings of the 13th International Conference on Pattern Recognition (ICPR), Vienna, Austria, 25–29 August 1996; pp. 350–358.
84. Rodriguez, Y.; Marcel, S. Face authentication using adapted local binary pattern histograms. In Proceedings of the 9th European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006; pp. 321–332.
85. Sadeghi, M.; Kittler, J.; Kostin, A.; Messer, K. A comparative study of automatic face verification algorithms on the banca database. In Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), Guilford, UK, 9–11 June 2003; pp. 35–43.
86. Huang, X.; Li, S.Z.; Wang, Y. Jensen-shannon boosting learning for object recognition. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; pp. 144–149.
87. Boutella, E.; Harizi, F.; Bengherabi, M.; Ait-Aoudia, S.; Hadid, A. Face verification using local binary patterns and generic model adaptation. *Int. J. Biomed.* **2015**, *7*, 31–44. [CrossRef]
88. Benzaoui, A.; Boukrouche, A. 1DLBP and PCA for face recognition. In Proceedings of the 2013 11th International Symposium on Programming and Systems (ISPS), Algiers, Algeria, 22–24 April 2013; pp. 7–11.
89. Benzaoui, A.; Boukrouche, A. Face Recognition using 1DLBP Texture Analysis. In Proceedings of the 5th International Conference of Future Computational Technologies and Applications, Valencia, Spain, 27 May–1 June 2013; pp. 14–19.
90. Benzaoui, A.; Boukrouche, A. Face Analysis, Description, and Recognition using Improved Local Binary Patterns in One Dimensional Space. *J. Control Eng. Appl. Inform. (CEAI)* **2014**, *16*, 52–60.
91. Ahonen, T.; Rathu, E.; Ojansivu, V.; Heikkilä, J. Recognition of Blurred Faces Using Local Phase Quantization. In Proceedings of the 19th International Conference on Pattern Recognition (ICPR), Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
92. Ojansivu, V.; Heikkil, J. Blur insensitive texture classification using local phase quantization. In Proceedings of the 3rd International Conference on Image and Signal Processing (ICSIP), Cherbourg-Octeville, France, 1–3 July 2008; pp. 236–243.
93. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In Proceedings of the 3rd International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), Rio de Janeiro, Brazil, 20 October 2007; pp. 168–182.
94. Lei, Z.; Ahonen, T.; Pietikainen, M.; Li, S.Z. Local Frequency Descriptor for Low-Resolution Face Recognition. In Proceedings of the 9th Conference on Automatic Face and Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 161–166.
95. Kannala, J.; Rahtu, E. BSIF: Binarized statistical image features. In Proceedings of the 21th International Conference on Pattern Recognition (ICPR), Tsukuba, Japan, 11–15 November 2012; pp. 1363–1366.
96. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
97. Deng, L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans. Signal Inf. Process.* **2014**, *3*, 1–29. [CrossRef]
98. Deng, L.; Yu, D. Deep Learning: Methods and Applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]

99. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
100. Salakhutdinov, R.; Hinton, G. Deep Boltzmann machines. In Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Clearwater, FL, USA, 16–19 April 2009; pp. 448–455.
101. Sutskever, I.; Martens, J.; Hinton, G. Generating text with recurrent neural networks. In Proceedings of the 28th International Conference on Machine Learning (ICML), Bellevue, WA, USA, 28 June–2 July 2011; pp. 1017–1024.
102. Poon, H.; Domingos, P. Sum-product networks: A new deep architecture. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 689–690.
103. Kimb, K.; Aminantoa, M.E. Deep Learning in Intrusion Detection Perspective: Overview and further Challenges. In Proceedings of the International Workshop on Big Data and Information Security (IWBIS), Jakarta, Indonesia, 23–24 September 2017; pp. 5–10.
104. Ouahabi, A. Analyse spectrale paramétrique de signaux lacunaires. *Traitement Signal* **1992**, *9*, 181–191.
105. Ouahabi, A.; Lacoume, J.-L. New results in spectral estimation of decimated processes. *IEEE Electron. Lett.* **1991**, *27*, 1430–1432. [CrossRef]
106. Scherer, D.; Müller, A.; Behnke, S. Evaluation of pooling operations in convolutional architectures for object recognition. In Proceedings of the 2010 International Conference on Artificial Neural Networks, Thessaloniki, Greece, 15–18 September 2010; pp. 92–101.
107. Coşkun, M.; Uçar, A.; Yildirim, Ö.; Demir, Y. Face recognition based on convolutional neural network. In Proceedings of the 2017 International Conference on Modern Electrical and Energy Systems (MEES), Kremenchuk, Ukraine, 15–17 November 2017; pp. 376–379.
108. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
109. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
110. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
111. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 2nd International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
112. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
113. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
114. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **2019**, *42*, 7132–7141.
115. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 539–546.
116. Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predicting 10,000 classes. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
117. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 1988–1996.

118. Sun, Y.; Wang, X.; Tang, X. Deeply learned face representations are sparse, selective, and robust. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2892–2900.
119. Sun, Y.; Liang, D.; Wang, X.; Tang, X. DeepID3: Face Recognition with Very Deep Neural Networks. *arXiv* **2015**, arXiv:1502.00873v1.
120. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Web-Scale training for face identification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2746–2754.
121. Ouahabi, A.; Depollier, C.; Simon, L.; Kouame, D. Spectrum estimation from randomly sampled velocity data [LDV]. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 1005–1012. [CrossRef]
122. Liu, J.; Deng, Y.; Bai, T.; Huang, C. Targeting ultimate accuracy: Face recognition via deep embedding. *arXiv* **2015**, arXiv:1506.07310v4.
123. Masi, I.; Tran, A.T.; Hassner, T.; Leksut, J.T.; Medioni, G. Do we really need to collect millions of faces for effective face recognition? In Proceedings of the 2016 European Conference on Computer Vision (ECCV), Amsterdam, The Netherland, 8–16 October 2016; pp. 579–596.
124. Zhang, X.; Fang, Z.; Wen, Y.; Li, Z.; Qiao, Y. Range loss for deep face recognition with Long-Tailed Training Data. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5419–5428.
125. Liu, W.; Wen, Y.; Yu, Z.; Yang, M. Large-margin softmax loss for convolutional neural networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 507–516.
126. Chen, B.; Deng, W.; Du, J. Noisy Softmax: Improving the Generalization Ability of DCNN via Postponing the Early Softmax Saturation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4021–4030.
127. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
128. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. *arXiv* **2013**, arXiv:1311.2901v3.
129. Ben Fredj, H.; Bouguezzi, S.; Souani, C. Face recognition in unconstrained environment with CNN. *Vis. Comput.* **2020**, 1–10. [CrossRef]
130. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 499–515.
131. Wu, Y.; Liu, H.; Li, J.; Fu, Y. Deep Face Recognition with Center Invariant Loss. In Proceedings of the Thematic Workshop of ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 408–414.
132. Yin, X.; Yu, X.; Sohn, K.; Liu, X.; Chandraker, M. Feature Transfer Learning for Face Recognition with Under-Represented Data. In Proceedings of the 2019 International Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
133. Ranjan, R.; Castillo, C.D.; Chellappa, R. L2-constrained softmax loss for discriminative face verification. *arXiv* **2017**, arXiv:1703.09507v3.
134. Deng, J.; Zhou, Y.; Zafeiriou, S. Marginal Loss for Deep Face Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2006–2014.
135. Wang, F.; Xiang, X.; Cheng, J.; Yuille, A.L. NormFace: L2 Hypersphere Embedding for Face Verification. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1041–1049.
136. Liu, Y.; Li, H.; Wang, X. Rethinking Feature Discrimination and Polymerization for Large-Scale Recognition. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), (Deep Learning Workshop), Long Beach, CA, USA, 4–9 December 2017.
137. Hasnat, M.; Bohné, J.; Milgram, J.; Gentric, S.; Chen, L. Von Mises-Fisher Mixture Model-based Deep Learning: Application to Face Verification. *arXiv* **2017**, arXiv:1706.04264v2.

138. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6738–6746.
139. Zheng, Y.; Pal, D.K.; Savvides, M. Ring Loss: Convex Feature Normalization for Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5089–5097.
140. Guo, Y.; Zhang, L. One-Shot Face Recognition by Promoting Underrepresented Classes. *arXiv* **2018**, arXiv:1707.05574v2.
141. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5265–5274.
142. Wang, F.; Cheng, J.; Liu, W.; Liu, H. Additive Margin Softmax for Face Verification. *IEEE Signal Process. Lett.* **2018**, *25*, 926–930. [CrossRef]
143. Wu, X.; He, R.; Sun, Z.; Tan, T. A Light CNN for Deep Face Representation with Noisy Labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [CrossRef]
144. Hayat, M.; Khan, S.H.; Zamir, W.; Shen, J.; Shao, L. Gaussian Affinity for Max-margin Class Imbalanced Learning. In Proceedings of the 2019 International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
145. Deng, J.; Guo, J.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the 2019 International Conference on Computer Vision and Pattern Recognition (CVPR), Lone Beach, CA, USA, 16–20 June 2019; pp. 4690–4699.
146. Huang, C.; Li, Y.; Loy, C.C.; Tang, X. Deep Imbalanced Learning for Face Recognition and Attribute Prediction. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **2019**. Available online: https://ieeexplore.ieee.org/document/8708977 (accessed on 21 July 2020).
147. Song, L.; Gong, D.; Li, Z.; Liu, C.; Liu, W. Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network. In Proceedings of the 2019 International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
148. Wei, X.; Wang, H.; Scotney, B.; Wan, H. Minimum margin loss for deep face recognition. *Pattern Recognit.* **2020**, *97*, 107012. [CrossRef]
149. Sun, J.; Yang, W.; Gao, R.; Xue, J.H.; Liao, Q. Inter-class angular margin loss for face recognition. *Signal Process. Image Commun.* **2020**, *80*, 115636. [CrossRef]
150. Wu, Y.; Wu, Y.; Wu, R.; Gong, Y.; Lv, K.; Chen, K.; Liang, D.; Hu, X.; Liu, X.; Yan, J. Rotation consistent margin loss for efficient low-bit face recognition. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–18 June 2020; pp. 6866–6876.
151. Ling, H.; Wu, J.; Huang, J.; Li, P. Attention-based convolutional neural network for deep face recognition. *Multimed. Tools Appl.* **2020**, *79*, 5595–5616. [CrossRef]
152. Wu, B.; Wu, H. Angular Discriminative Deep Feature Learning for Face Verification. In Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 2133–2137.
153. Chen, D.; Cao, X.; Wang, L.; Wen, F.; Sun, J. Bayesian face revisited: A joint formulation. In Proceedings of the European Conference on Computer Vision (ECCV), Firenze, Italy, 7–13 October 2012; pp. 566–579.
154. Chen, B.C.; Chen, C.S.; Hsu, W.H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans. Multimed.* **2015**, *17*, 804–815. [CrossRef]
155. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 3730–3738.
156. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
157. Oumane, A.; Belahcene, M.; Benakcha, A.; Bourennane, S.; Taleb-Ahmed, A. Robust Multimodal 2D and 3D Face Authentication using Local Feature Fusion. *Signal Image Video Process.* **2016**, *10*, 12–137. [CrossRef]
158. Oumane, A.; Boutella, E.; Benghherabi, M.; Taleb-Ahmed, A.; Hadid, A. A Novel Statistical and Multiscale Local Binary Feature for 2D and 3D Face Verification. *Comput. Electr. Eng.* **2017**, *62*, 68–80. [CrossRef]

159. Soltanpour, S.; Boufama, B.; Wu, Q.M.J. A survey of local feature methods for 3D face recognition. *Pattern Recognit.* **2017**, *72*, 391–406. [CrossRef]
160. Zhou, S.; Xiao, S. 3D Face Recognition: A Survey. *Hum. Cent. Comput. Inf. Sci.* **2018**, *8*, 8–35. [CrossRef]
161. Min, R.; Kose, N.; Dugelay, J. KinectFaceDB: A Kinect Database for Face Recognition. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 1534–1548. [CrossRef]
162. Drira, H.; Ben Amor, B.; Srivastava, A.; Daoudi, M.; Slama, R. 3D Face Recognition under Expressions, Occlusions, and Pose Variations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2270–2283. [CrossRef] [PubMed]
163. Ribeiro Alexandre, G.; Marques Soares, J.; Pereira Thé, G.A. Systematic review of 3D facial expression recognition methods. *Pattern Recognit.* **2020**, *100*, 107108. [CrossRef]
164. Ríos-Sánchez, B.; Costa-da-Silva, D.; Martín-Yuste, N.; Sánchez-Ávila, C. Deep Learning for Facial Recognition on Single Sample per Person Scenarios with Varied Capturing Conditions. *Appl. Sci.* **2019**, *9*, 5474.
165. Kim, D.; Hernandez, M.; Choi, J.; Medioni, G. Deep 3D face identification. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 133–142.
166. Gilani, S.Z.; Mian, A.; Eastwood, P. Deep, dense and accurate 3D face correspondence for generating population specific deformable models. *Pattern Recognit.* **2017**, *69*, 238–250. [CrossRef]
167. Gilani, S.Z.; Mian, A.; Shafait, F.; Reid, I. Dense 3D face correspondence. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **2018**, *40*, 1584–1598. [CrossRef] [PubMed]
168. Gilani, S.Z.; Mian, A. Learning from Millions of 3D Scans for Large-scale 3D Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1896–1905.
169. Mimouna, A.; Alouani, I.; Ben Khalifa, A.; El Hillali, Y.; Taleb-Ahmed, A.; Menhaj, A.; Ouahabi, A.; Ben Amara, N.E. OLIMP: A Heterogeneous Multimodal Dataset for Advanced Environment Perception. *Electronics* **2020**, *9*, 560. [CrossRef]
170. Benzaoui, A.; Boukrouche, A.; Doghmane, H.; Bourouba, H. Face recognition using 1DLBP, DWT, and SVM. In Proceedings of the 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), Tlemcen, Algeria, 25–27 May 2015; pp. 1–6.
171. Ait Aouit, D.; Ouahabi, A. Monitoring crack growth using thermography.-Suivi de fissuration de matériaux par thermographie. *C. R. Mécanique* **2008**, *336*, 677–683. [CrossRef]
172. Arya, S.; Pratap, N.; Bhatia, K. Future of Face Recognition: A Review. *Procedia Comput. Sci.* **2015**, *58*, 578–585. [CrossRef]
173. Zafeiriou, S.; Zhang, C.; Zhang, Z. A survey on face detection in the wild: Past, present and future. *Comput. Vis. Image Underst.* **2015**, *138*, 1–24. [CrossRef]
174. Min, R.; Xu, S.; Cui, Z. Single-Sample Face Recognition Based on Feature Expansion. *IEEE Access* **2019**, *7*, 45219–45229. [CrossRef]
175. Zhang, D.; An, P.; Zhang, H. Application of robust face recognition in video surveillance systems. *Optoelectron. Lett.* **2018**, *14*, 152–155. [CrossRef]
176. Tome, P.; Vera-Rodriguez, R.; Fierrez, J.; Ortega-Garcia, J. Facial soft biometric features for forensic face recognition. *Forensic Sci. Int.* **2015**, *257*, 271–284. [CrossRef] [PubMed]
177. Fathy, M.E.; Patel, V.M.; Chellappa, R. Face-based Active Authentication on mobile devices. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia, 19–24 April 2015; pp. 1687–1691.
178. Medapati, P.K.; Murthy, P.H.S.T.; Sridhar, K.P. LAMSTAR: For IoT-based face recognition system to manage the safety factor in smart cities. *Trans. Emerg. Telecommun. Technol.* **2019**, 1–15. Available online: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.3843?af=R (accessed on 10 July 2020).