Computational Visual Media
https://doi.org/10.1007/s41095-022-0317-1 Vol. 9, No. 4, December 2023, 657–685

Review Article

3D face recognition: A comprehensive survey in 2022

Yaping Jing[1], Xuequan Lu[1] (✉), and Shang Gao[1]

Abstract In the past ten years, research on face recognition has shifted to using 3D facial surfaces, as 3D geometric information provides more discriminative features. This comprehensive survey reviews 3D face recognition techniques developed in the past decade, both conventional methods and deep learning methods. These methods are evaluated with detailed descriptions of selected representative works. Their advantages and disadvantages are summarized in terms of accuracy, complexity, and robustness to facial variations (expression, pose, occlusion, etc.). A review of 3D face databases is also provided, and a discussion of future research challenges and directions of the topic.

Keywords 3D face recognition; 3D face databases; deep learning; local features; global features

1 Introduction

Face recognition has become a commonly used biometric technology. It is widely applied in public surveillance, authentication, security, intelligence, and many other systems [1]. During recent decades, many 2D face recognition techniques have achieved strong results in controlled environments. The accuracy of 2D face recognition has been greatly enhanced especially by the emergence of deep learning. However, these techniques are still challenged by the intrinsic limitations of 2D images, due to variations in illumination, pose, and expression, occlusion, disguises, time delays, and image quality [2]. 3D face recognition can outperform 2D face recognition [3] with greater recognition accuracy and robustness, as it is less sensitive to pose, illumination, and expression [4]. Thus, 3D face recognition has become an active research topic in recent years.

Face recognition mainly involves the process of extracting feature representations from the input face, matching the extracted features with existing databases and predicting the personal identity of the input face. Therefore, using rich facial features is critical to the recognition result. In 3D face recognition, 3D face data are used for training and testing purposes. Compared to 2D images, 3D faces contain richer geometric information, that can provide more discriminative features and help face recognition systems overcome the inherent defects and drawbacks of 2D face recognition, such as facial expression, occlusion, and pose variation. Furthermore, 3D data are relatively unchanged after scaling, rotation, and illumination change [5]. Most 3D scanners can acquire both 3D meshes/point clouds and corresponding textures. This allows us to integrate advanced 2D face recognition algorithms into 3D face recognition systems for better results.

One of the main challenges to 3D face recognition is the acquisition of 3D training images—this cannot be accomplished by crawling the Web, unlike for 2D face images. It requires special hardware instead. According to the technologies used, collection systems can be broadly divided into active acquisition and passive acquisition [6]. An active collection system actively emits invisible light (e.g., an infrared laser beam) to illuminate the target face and obtain the shape features of the target by measuring reflectivity. A passive acquisition system consists of several cameras placed apart. It matches points observed to those from other cameras and calculates the 3D position of the matched point. The 3D surface is formed from a set of matched points.

Since 2000, many researchers have begun to assess 3D face recognition algorithms on large-scale

1 The School of Information Technology, Deakin University, Waurn Ponds, VIC, Australia. E-mail: Y. Jing, jingyap@deakin.edu.au; X. Lu, xuequan.lu@deakin.edu.au (✉); S. Gao, shang.gao@deakin.edu.au.

Manuscript received: 2022-04-14; accepted: 2022-09-29

databases and published related 3D face databases, e.g., Face Recognition Vendor Tests (FRVT-2000) [7], FRVT-2002 [8], the Face Recognition Grand Challenge (FRGC) [9], and FRVT-2006 [10]. This suggests that there is a close relationship between large datasets and 3D face recognition techniques, so we also summarize existing public 3D face databases and their data augmentation methods as well as reviewing recognition technologies.

Several relevant surveys have been conducted by researchers from different perspectives. In 2006, Ref. [3] reviewed research trends in 3D face recognition. Ref. [11] summarized the associated literature up to the year 2007. Ref. [12] studied various algorithms for expression invariant 3D face recognition and evaluated the complexity of existing 3D face databases. Later, Ref. [2] categorized face recognition algorithms into single-modal and multi-modal ones in 2014. Ref. [1] studied 3D face recognition techniques, comprehensively covering conventional methods. Recently, Refs. [13] and [6] each presented a review of 3D face recognition algorithms, but only a few deep learning-based methods were covered. Refs. [14] and [15] reviewed deep learning-based face recognition methods in 2018, but their focus was mainly on 2D face recognition.

In this paper, we focus on 3D face recognition. Compared to the existing literature, the main contributions of our work are follows:

- this is the first survey paper to comprehensively cover conventional methods and deep learning-based methods for 3D face recognition;
- unlike existing surveys, it pays special attention to deep learning-based 3D face recognition methods;
- it covers the latest and most advanced developments in 3D face recognition, providing a clear progress chart for the topic; and
- it provides a comprehensive comparison of existing methods using available databases, and suggests future research challenges and directions.

According to the feature extraction methods adopted, 3D face recognition techniques can be divided into two categories: conventional methods and deep learning-based methods (see Fig. 1). To extract face features, conventional methods always use traditional algorithms, either linear or nonlinear, e.g., principle component analysis (PCA). They can be further divided into three types: local, global, and hybrid. As for deep learning-based methods, nearly all use pre-trained networks and then fine-tune these networks with converted data (e.g., 2D maps from 3D faces). Popular deep learning-based face recognition networks include VGGNet [16], ResNet [17], ANN [18], and recent lightweight CNNs such as MobileNetV2 [19].

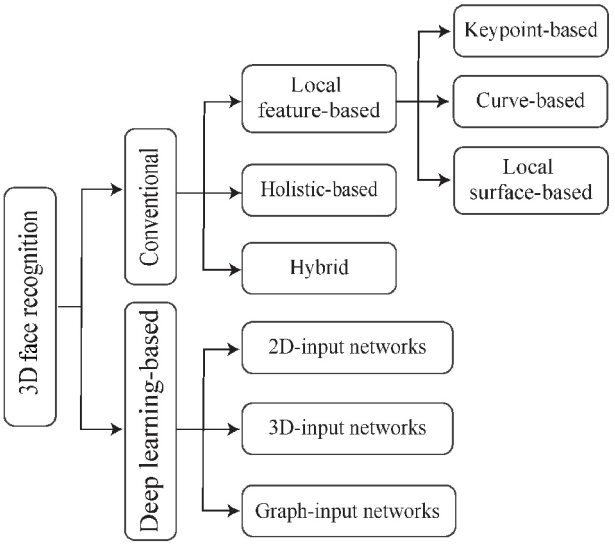


Fig. 1 Taxonomy of 3D face recognition methods.

The structure of this survey is as follows. Section 2 introduces widely used 3D face databases and datasets. Section 3 covers data preprocessing and augmentation. Sections 4 and 5 respectively review conventional 3D face recognition methods and deep learning-based methods. Section 6 compares these methods and discusses future research directions, followed by conclusions in Section 7.

2 3D face databases

Large-scale 3D face databases and datasets are essential for the development of 3D face recognition. They are used to train feature extraction algorithms and evaluate their output. To meet this demand, many research institutions and researchers have created various 3D face databases. Table 1 lists current prominent 3D face databases and compares their data formats, number of persons contained (IDs), image variations (e.g., expression, pose, and occlusion), and scanning devices. Four different 3D data formats are in use: point clouds (Fig. 2(a)), 3D meshes (Fig. 2(b)), range images (Fig. 2(c)), and depth maps plus 3D video.

Table 1 3D face databases. Datatype: M=mesh, P=point cloud, R=range image, V=video, 3DV=3D video; Expression: S=smile/happiness, M=multiple expressions; Pose: LR=slight left/right turn, UD=slight up/down turn, M=multiple poses

	Year	Data type	IDs	Scans	Texture	Expression	Pose	Occlusion	Scanner
3DRMA [21]	2000	M	120	720	Yes	—	LR, UD	—	Standard CCD black and white camera
FSU [22]	2003	M	37	222	No	—	—	—	Minolta Vivid 700
GavabDB [23]	2004	M	61	427	No	S	±30°	—	Minolta Vi-700 laser range scanner
FRGC v2 [9]	2005	R	466	4007	Yes	S	±15°	—	Minolta Vivid 3D scanner
UND [24]	2005	R	275	670	Yes	—	±45°, ±60°	—	Minolta Vivid 900
ZJU-3DFED [25]	2006	M	40	360	No	S, surprise, sad	—	—	InSpeck 3D MEGA Capturor DF
BU3D-FE [26]	2006	M	100	2500	Yes	M	—	—	Stereo photography, 3DMD digitizer
CASIA [27]	2006	R	123	4059	No	M	±90°	—	Minolta Vivid 910
FRAV3D [28]	2006	M	105	1696	Yes	M	M	—	Minolta Vivid 700 red laser light scanner
ND-2006 [29]	2007	R	888	13,450	Yes	M	±15°	—	Minolta Vivid 910
Bosphorus [30]	2008	P	105	4666	Yes	M	M	4	Inspeck Mega Capturor II 3D scanner
UoY [31]	2008	M	350	5000	Yes	M	UD	—	Stereo vision 3D camera
SHREC08 [32]	2008	R	61	427	No	M	UD	—	—
BJUT-3D [33]	2009	M	500	1200	Yes	—	—	—	Cyberware 3D laser scanner
Texas-3D [34]	2010	R	118	1149	Yes	M	±10°	—	MU-2 stereo imaging system
UMBDB [20]	2011	R	143	1473	Yes	S, anger, bored	—	7	Minolta Vivid 900 laser scanner
3D-TEC [35]	2011	R	214	428	Yes	S	—	—	Minolta scanner
SHREC11 [36]	2011	R	130	780	No	—	M	—	Escan laser scanner
NPU3D [37]	2012	M	300	10,500	No	M	M	4	Konica Minolta Vivid 910 non-contact 3D laser scanner
BU4D-FE [38]	2013	3DV	101	60,600	Yes	—	—	—	Di3D (Dimensional Imaging) dynamic system
KinectFaceDB [39]	2014	R	52	936	Yes	N, S, surprise	LR	Multiple	Kinect
Lock3DFace [40]	2016	R	509	5711	Yes	M	±90°	Random cover	Kinect
F3D-FD [41]	2018	R	2476	—	Yes	—	Semi-lateral with ear	Half face	Vectra M1 scanner
LS3DFace [42]	2018	P	1853	31,860	Yes	—	—	—	—
4DFAB [43]	2018	V	180	1800k+	Yes	M	M	—	DI4D capturing system
WFFD [44]	2020	V	241	285	Yes	—	—	—	—
SIAT-3DFE [45]	2020	3D	500	8000	Yes	M	—	2	Structured light (CASZM-3D)
FaceScape [46]	2020	Y	938	18,760	Yes	M	—	—	68 DSLR cameras
CAS-AIR-3D Face [47]	2021	V	3093	24,713	No	S, surprise	±90°	Glasses	Intel RealSense SR305

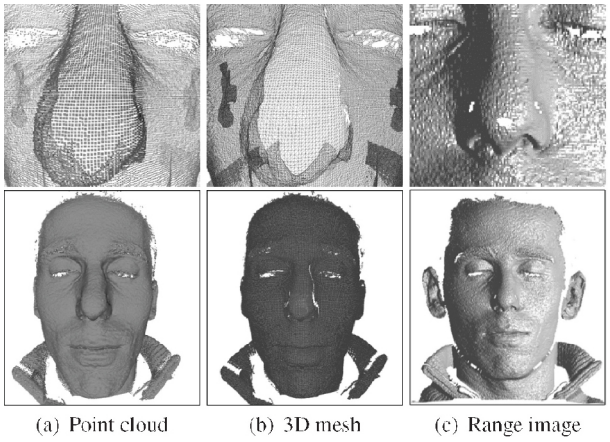


Fig. 2 3D face data representations. Reproduced with permission from Ref. [20], © IEEE 2011.

Before 2004, there were few public 3D face databases. Some representatives include *3DRMA* [21], *FSU* [22], and *GavabDB* [23]. The *GavabDB* database contains 61 individuals, aged between 18 and 40. Each has 3 frontal images with different expressions and 4 rotated images with a neutral expression [23]. In 2005, the *FRGC v2* database was designed to improve the results of face recognition algorithms. It had a huge impact on the development of 3D face recognition [9], and it is still used as a standard reference database for evaluating 3D face recognition algorithms. In the same year, another important database, the University of Notre Dame *UND* database [24] was released, in which each person has only one 3D image and multiple 2D images [24].

From 2006 to 2010, further databases were created. The largest is *ND-2006* [29] which is a superset of FRGC v2. It contains 13,450 images and 888 persons with as many as 63 images per person [29]. The second largest is *UoY* [31], which consists of more than 5000 models (350 persons) and is owned by the University of York (UK) [31]. *CASIA* [27] and *Bosphorus* [30] are similar in size, with close to 5000 images. *CASIA* was collected in 2004 using a non-contact 3D digitizer, a Minolta Vivid 910, and contains 4059 images of 123 subjects [27]. It not only considers separate variations in expression, pose, and illumination, but also introduces combined changes, with different expressions in different poses. *Bosphorus* has 381 individuals and the most expression and pose changes. It provides manual marking of 24 facial landmarks for each scanned image, including nose tip, chin centre, and eye corners [30]. Another database with manual landmarks is *Texas-3D* [34], whose 3D images have been preprocessed, and 25 manual landmarks added. Therefore, it provides a good option for researchers to focus specifically on developing 3D face recognition algorithms, without considering initial preprocessing of 3D images [34].

BU3D-FE [26] (Binghamton University 3D Facial Expression) is a database specifically developed for 3D facial expression recognition. It contains 100 identities with 6 expression types: anger, happiness, sadness, surprise, disgust, and fear [26]. The *FRAV3D* [28] database involves 81 males and 24 females; three kinds of images (3D meshes, 2.5D range data, and 2D color images) were captured using the MINOLTA VIVID-700 red laser scanner [28]. *BJUT-3D* [33] is one of the largest Chinese 3D face databases and includes 1200 Chinese 3D face images [33]. The two smallest databases are *ZJU-3DFED* [25] and *SHREC08* [32]. *ZJU-3DFED* consists of 40 identities and 9 scans with four different expressions for each identity [25]. *SHREC08* consists of 61 people with 7 scans for each [32].

Between 2010 and 2015, there were six noteworthy databases created. *UMBDB* [20] is an excellent database for testing 3D face recognition algorithms with varying occlusion. It contains 578 occluded acquisitions [20]; some examples are shown in Fig. 3. *3D-TEC* [35] (3D Twins Expression Challenge) is a challenging dataset as it contains 107 pairs of twins with similar faces and different expressions [35]. Thus, this database is helpful when 3D face recognition must work in the presence of varying expressions.

Fig. 3 Examples from the UMB database. Row 1: neutral faces under differing illumination. Rows 2, 3: faces partially occluded by objects such as scarves and hands. Reproduced with permission from Ref. [20], © IEEE 2011.

SHREC11 [36] is built upon a new collection of 130 masks with 6 3D face scans [36]. Like *BJUT-3D*, Northwestern Polytechnic University 3D (*NPU3D* [37]) is another large-scale Chinese 3D face database, composed of 10,500 3D face captures, corresponding to 300 individuals [37]. *BU4D-FE* [38] is a 3D video database that records spontaneous expressions of various young people by completing 8 emotional expression elicitation tasks [38]. *KinectFaceDB* [39] was the first publicly available face database based on the Kinect sensor and contains four data modalities (2D, 2.5D, 3D, and video-based) [39].

Recently, another large-scale 3D face database *Lock3DFace* [40] was released. It is based on Kinect and contains several variations in expression, pose, time-lapse, and occlusion [40]. *F3D-FD* [41] is a large dataset, and has the most individuals: 2476. For each individual, it includes partial 3D scans from frontal and two semi-lateral views, and a one-piece face with lateral parts (including ears and earless, with landmarks) [41]. *LS3DFace* [42] is the largest dataset so far, including 31,860 3D face scans of 1853 people. It is combines data from multiple challenging public datasets, including FRGC v2, BU3D-FE, Bosphorus, GavabDB, Texas-3D, BU4D-FE, CASIA, UMBDB, 3D-TEC, and ND-2006 [42]. *4DFAB* [43] is a large dynamic high-resolution 3D face database; it contains 4D videos of subjects showing spontaneous and posed facial behavior.

The large-scale Wax Figure Face Database (*WFFD* [44]) is designed to address vulnerabilities in existing 3D facial spoofing databases and to promote the research of 3D facial presentation attack detection [44]. This database includes photo-based and video-based data. We only detail the video information in Table 1. *SIAT-3DFE* [45] is a 3D facial expression dataset in which every identity has 16 facial expressions including natural, happy, sad, surprised, and several exaggerated expressions (open mouth, frowning, etc.), as well as two occluded 3D cases [45]. Another recent database is *FaceScape* [46], which consists of 18,760 textured 3D data with pore-level facial geometry [46]. *CAS-AIR-3D* [47] is a large low-quality 3D face database, including 3093 individuals. Each records 8 videos with different poses and expressions, occlusion, and distance change.

It is well known that the performance of 3D face recognition algorithms may vary on different 3D face databases. Increasing the gallery size may degrade the performance of face recognition [48]. Although some algorithms have achieved good results on these existing 3D face databases, they still cannot be used in the real world due to less controlled conditions. The establishment of large-scale 3D face databases to simulate real-world situations is essential to facilitate research into 3D face recognition. In addition, collecting 3D face data is a time-consuming and resource-demanding task. Research into large dataset generation algorithms is one of our suggestions for future work (see also Section 3.2).

3 Data preprocessing and augmentation

3.1 Data preprocessing

In most situations, the acquired raw 3D face data cannot be directly input to feature extraction systems as they may contain redundant information [6]. For example, presence of hair, neck, and background may affect the accuracy of recognition. Thus, 3D data are usually preprocessed before being passed into a feature extraction model.

In general, the data preprocessing phase includes three main steps: facial landmark detection and orientation, data segmentation, and face registration. Facial landmarks are a set of keypoints defined by anthropometric studies [49] and can be used to automatically localize and register a face. Some databases already provide landmarks for each face image. Data segmentation is the process of utilizing facial landmarks, such as nose tip and eye corners, to segment the facial surface [49]. This process is always used by local conventional methods, which determine identifiable facial parts like the nose and eyes for feature extraction. As an essential step before feature extraction and matching, face registration aligns the target surface (the entire face or face part) with the training surface in the gallery.

3.2 Data augmentation

To improve the performance and robustness of face recognition systems, large-scale datasets are required, especially for deep learning-based methods, since their networks need to be trained using a large amount of training data.

Several augmentation methods can be used to increase the size of the training and test datasets.

The easiest is to rotate and crop existing face data. Another popular approach is to use 3D morphable facial models (3DMM) [50] to generate new shapes and expressions to synthesize new facial data [51, 52]. Randomly selecting sub-feature sets from different samples of a person and combining them to generate a new face is also a reliable way to enrich the identities in datasets [42].

Recently, generative adversarial networks (GANs) have been used to generate realistic synthetic images [53–58]. GANs usually consist of a generator and a discriminator, which are alternately trained in minimax games. The discriminator is trained to discriminate generated samples from real samples, and the generator is trained to generate images resembling real ones to minimize the success of the discriminator.

Ref. [53] proposed UV-GAN, which first generates a completed u–v map from a single image, then attaches the completed u–v to a fitted 3D mesh, and generates synthetic faces in arbitrary poses to increase pose variation. In 3D-PIM [54], a 3DMM cooperates with a GAN to automatically recover natural frontal face images from arbitrary poses. The 3DMM is used as a simulator to generate synthetic faces with normalized poses, and the GAN is used to refine the realism of the output detail.

FaceID-GAN [55] formulates a three-player GAN by introducing an identity classifier that works with the discriminator to compete with the generator. Ref. [56] also proposed a method to generate frontal faces from profile faces by employing a GAN with dual-discriminator structure. The 3D GAN generator in Ref. [57] is augmented with an integrated 3DMM, including two CNNs for facial texture and background generation, to ensure that identity is preserved in the synthetic images after manipulating pose.

In FA-GAN [58], a graph-based two-stage architecture was proposed, consisting of two parts: a geometry preservation module (GPM) and a face disentanglement module (FDM). For the GPM, a graph convolutional network (GCN) [59] is introduced to explore the relationship between different face regions and better preservation of geometric information. FDM disentangles the encoded facial feature embeddings into identity representations and deformation attribute codes (pose, expression), which can be further used for face manipulation and enhancement.

4 Conventional methods

4.1 Approaches

According to Ref. [60] and our review of the last decade's literature, conventional face recognition algorithms can be classified into three types based on their feature extraction approaches, local, global, and hybrid, as shown in Fig. 1. Local approaches mainly focus on local facial features such as the nose and eyes [60]. In contrast to local methods, global approaches use the entire face to generate feature vectors for feature classification. Hybrid methods use both local and global facial features.

In local methods, fusion schemes are used to improve accuracy. There are five levels for fusion schemes: sensor level, feature level, rank level, decision level, and score level [1]. Sensor level fusion merges the original sensor data at the initial stage of recognition; feature level fusion involves the combination of features extracted from different facial representations of one single object. For rank level fusion, ranks are assigned to gallery images based on a descending sequence of confidence, while score level fusion combines matching scores of each classifier based on a weighting scheme. Decision level fusion combines the decision of each classifier [1].

Details of the three conventional face recognition approaches are discussed below.

4.2 Local methods

4.2.1 Classification

In the last decade, many local approaches have been built, where local feature descriptors are used to describe 3D local facial information. Table 2 lists noteworthy 3D local methods and summarizes their important details.

Following Ref. [13], these methods can be classified into three different types based on the descriptors: keypoint-based, curve-based, and region-based. Keypoint-based methods detect a set of 3D keypoints based on face geometry and build feature descriptors by calculating relationships between these keypoints. Curve-based methods use a set of curves on each face surface as feature vectors. Region-based methods extract features from certain regions of the face surface [13].

Table 2 Local techniques. RR1 = rank-1 recognition rate. Advantages: (i) changes to which the method is relatively robust, or (ii) other benefits such as speed, etc. Limitations note circumstances under which the corresponding method under-performs

Author/year	Category	Method	Advantage	Limitation	Database	RR1 (%)
Berretti et al. (2011) [61]	SIFT keypoint	Covariance matrix, X^2 dist	Partial facial	Keypoints redundancy	FRGC v2	89.2 (partial faces)
Li et al. (2011) [62]	Mesh-based keypoint	Histograms, cosine dist	Expression	Pose	Bosphorus	94.1
Creusot et al. (2011) [63]	Landmark keypoint	Linear combination	Expression	Computationally expensive	FRGC v2	—
Zhang and Wang (2011) [64]	Landmarks keypoint	SVM-based fusion, six similarity measures	Simple preprocessing, noise, resolution	Occlusion	FRGC v2	96.2
Inan and Halici (2012) [65]	SIFT keypoint	Cosine dist	Neutral expression	Noise	FRGC v2	97.5
Berretti et al. (2013) [66]	Curve	Sparse	Missing parts	Large pose, expression	FRGC v2 GavabDB UND	95.6 97.13 75
Li and Da (2012) [67]	Curve	PCA	Expression, hair occlusion	Exaggerated expressions	FRGC v2	97.80
Ballihi et al. (2012) [68]	Curve	Euclidean dist, AdaBoost	Efficient, data storage	Occlusion	FRGC v2	98
Berretti et al. (2013) [69]	Mesh-based keypoint	X^2 dist	Missing parts	Low accuracy	UND	77.1
Smeets et al. (2013) [70]	Mesh-based keypoint	Angles comparison	Expression, partial data	Noise	Bosphorus FRGC v2	93.7 89.6
Creusot et al. (2013) [71]	Mesh-based landmark keypoint	Linear (LDA), non-linear (AdaBoost)	Expression	Complexity, occlusion	FRGC v2 Bosphorus	— —
Tang et al. (2013) [72]	Region (LBP-based)	LBP, Nearest-neighbor (NN)	Expression	Occlusion, missing data	FRGC v2	94.89
Lei et al. (2013) [73]	Region (geometric feature)	SVM	Expression	Occlusion	FRGC v2 BU-3DFE	95.6 97.7
Elaiwat et al. (2014) [74]	Region	Curvelet transform	Illumination, expression	Occlusion	FRGC v2	—
Drira et al. (2013) [75]	Curve	Riemannian framework	Pose, missing data	Extreme expression, complexity	FRGC v2	97.7
Li et al. (2014) [76]	Region (LBP-based)	ICP, sparse-based	Expression, fast	Pose, occlusion	FRGC v2	96.3
Berretti et al. (2014) [77]	Mesh-based keypoint	Classifier	Occlusion, missing parts	Noise, low-resolution images	Bosphorus	94.5
Lei et al. (2014) [78]	Curve	KPCA, SVM	Efficient, expression	Occlusion	FRGC v2 SHREC08	— —
Tabia et al. (2014) [79]	Region (Geometric features)	Riemannian metric	Expression	Occlusion	GavabDB	94.91
Vezzetti et al. (2014) [80]	Landmark keypoint	Euclidean distance	Expression, occlusion	Low accuracy	Bosphorus	—
Li et al. (2015) [81]	Mesh-based keypoint	Gaussian filters, fine-grained matcher	Expression, occlusion, registration-free	Cost	Bosphorus	96.56

Table 2 Local techniques (continued)

Author/year	Category	Method	Advantage	Limitation	Database	RR1 (%)
Elaiwat et al. (2015) [82]	Mesh-based keypoint	Curvelet transform, cosine dist	Illumination, expressions	Occlusion	FRGC v2	97.1
Al-Osaimi (2016) [83]	Curve	Euclidean dist	Fast, expression	Occlusion	FRGC	97.78
Ming (2015) [84]	Region	Regional, global regression	Large pose, efficient	Patches detection	FRGC v2 CASIA BU-3DFE	— — —
Guo et al. (2016) [85]	Keypoint	Rotational Projection Statistics (RoPS), average dist	Occlusion, expression and pose	Cost	FRGC v2	97
Soltanpour and Wu (2017) [86]	SIFT keypoint	Histogram matching	Expression	Pose	FRGC v2	96.9
Lei et al. (2016) [87]	SIFT keypoint	Two-Phase Weighted	Missing parts, occlusion, data corruptions	Extreme pose, expression	FRGC v2	96.3
Emambakhsh and Evans (2017) [88]	Curve	Mahalanobis, cosine dist	Expression, single sample	Occlusion	FRGC v2	97.9
Werghi et al. (2016) [89]	Region (LBP-based)	Cosine, X^2 dist	Expression, missing data	pose	BU-3DFE Bosphorus	— —
Hariri et al. (2016) [90]	Region (geometric features)	Geodesic dist	Expression, pose	Partial occlusion	FRGC v2	99.2
Soltanpour and Wu (2017) [91]	Region (LDP)	ICP	Expression	Extreme pose, missing data	FRGC v2 Bosphorus	98.1 97.3
Deng et al. (2017) [92]	SIFT keypoints	Riemannian kernel sparse coding	Low-complex	Expression, occlusion	FRGC v2	97.3
Abbad et al. (2018) [93]	Curve	Angles comparison	Expression, time consumption	Occlusion, missing data	GavabDB	99.18
Soltanpour and Wu (2019) [94]	Region (LDP)	ICP	Computational cost	Pose	FRGC v2	99.3
Shi et al. (2020) [95]	Region (LBP-based)	LBP, SVM	Low consumption	Pose, occlusion	Texas-3D	96.83

4.2.2 Keypoint-based methods

As keypoint-based methods use a set of keypoints and their geometric relationships to represent facial features, two important steps are involved: keypoint detection and feature descriptor construction [13].

One of the most commonly used keypoint detectors is the scale invariant feature transformation (SIFT) [98]. For example, Ref. [61] used SIFT to detect relevant keypoints of a 3D depth image (see Fig. 1 in Ref. [61]), where local shape descriptors are adopted to measure the changes of face depth in each keypoint's neighborhood. In Ref. [65], to obtain feature vectors, SIFT descriptors are applied to 2D matrices including shape index, curvedness, and Gaussian and mean curvature values generated from 3D face data.

Ref. [86] used SIFT keypoint detection on pyramidal shape maps to obtain 3D geometric information and combine it with 2D keypoints. However, this method is sensitive to pose changes. Ref. [85] used a 3D point cloud registration algorithm combined with local features to achieve both pose and expression invariance. Later, a keypoint-based multiple triangle statistics (KMTS) [87] method was proposed to address partial facial data, pose change, and large facial expression variation. Recently, SIFT has also been used to detect keypoints in Ref. [92]; it uses local covariance descriptors and Riemann kernel sparse coding to improve the accuracy of 3D face recognition. Accuracy was further improved in Ref. [99].

In order to improve the robustness to large

occlusion or pose variation, SIFT keypoint detection is directly used on 3D mesh data. The extension of SIFT to 3D meshes is called MeshSIFT [70]. In Ref. [70], salient points on a 3D face surface were first detected as extreme values in scale space, and then an orientation was assigned to these points. A feature vector was used to describe them by concatenating the histograms of slant angles and shape indices. Before this approach was applied, Ref. [62] also used minimum and maximum curvatures within a 3D Gaussian scale space to detect salient points and used the histograms of multiple order surface differential quantities to characterize the local facial surface. The descriptors of detected local regions were further used in 3D face local matching. Ref. [81] also described an extension to this work, in which a fine-grained matching of 3D keypoint descriptors was proposed to enlarge intra-subject similarity and reduce inter-subject similarity. However, a large number of keypoints were detected by these methods.

A meshDOG keypoint detector was proposed by Berretti et al. [69, 77]. They first used the meshDOG keypoint detector and local geometric histogram (GH) descriptor to extract features, and then selected the most effective feature based on an analysis of the optimal scale, distribution and clustering of keypoints, and the features of local descriptors. Recently, Ref. [82] exploited a curvelet-based multimodal keypoint detector and local surface descriptor that extracts both texture and 3D local features. It reduces the computational cost of keypoint detection and feature building, as the curvelet transform is based on the FFT.

Additionally, a set of facial landmarks is used for creating feature vectors in some methods, and a shape index is widely used to detect landmarks. In Ref. [63], keypoints were extracted from a shape dictionary, which was learned on a set of 14 manually placed landmarks on a human face. As an extension, Ref. [71] used a dictionary of learned local shapes to detect keypoints, and evaluated them through linear (LDA) and nonlinear (AdaBoost) methods. Ref. [64] detected resolution invariant keypoints and scale-space extremes on shape index images based on scale-space analysis, and used six scale-invariant similarity measures to calculate the matching score. In Ref. [80], an entirely geometry-based 3D face recognition method was proposed and 17 landmarks were automatically extracted based on facial geometrical characteristics; this was further extended in Ref. [100].

4.2.3 Curve-based methods

A curve-based method uses a set of curves to construct feature descriptors. It is difficult to decide whether such methods are local or global, because these curves usually cover the entire face, and capture geometric information from different face regions to represent the 3D face. The curves can be grouped into level curves and radial curves according to their distribution. Level curves are non-intersecting closed curves of different lengths; radial curves are open curves, usually starting from the nose tip.

Level curves can be further divided into iso-depth and iso-geodesic curves [13] (see Fig. 4 in Ref. [96] and Fig. 5 in Ref. [101]). Iso-depth curves can be obtained by translating a plane across the facial surface in one direction and were first introduced by Samir et al. [97]. Ref. [96] expanded this work and proposed iso-geodesic curves which are level curves of surface distance from the nose tip. However, both kinds are sensitive to occlusion, missing parts, and large facial expressions. Thus, radial curves were introduced in Ref. [101] and extended in Ref. [75]. These curves can better handle occlusion and missing parts as it is uncommon to lose a full radial curve and at least some parts of a radial curve can be used. Also, they can be associated with different facial expressions as the radial curves pass through different facial regions.

In Ref. [67], facial curves in the nose region of a target face were first extracted to form a rejection classifier, which was used to quickly and effectively eliminate different faces in the gallery. Then the face was segmented into six facial regions. A facial deformation mapping was produced by using curves in these regions. Finally, adaptive regions were selected to match the two identities. In Ref. [68], geometric curves from the level sets (circular curves) and streamlines (radial curves) through the Euclidean distance functions of 3D faces were combined for high-accuracy face recognition.

A highly compact signature of a 3D face can be characterized by a small set of features selected by the Adaboost algorithm [102], a well-known machine learning feature selection method. By using the

selected curves, face recognition time was reduced from 2.64 to 0.68 s, showing that feature selection can effectively improve system performance. To provide high discriminative feature vectors and improve computational efficiency, angular radial signatures (ARSs) were proposed by Lei et al. [78]. An ASR is a set of curves from the nose tip (the origin of facial range images) at intervals of θ radians.

Another type of facial curve was introduced by Berretti et al. [66]. SIFT was utilized to detect keypoints of 3D depth images and which were connected to form the facial curves. A 3D face can be represented by a set of facial curves built from matched keypoints. Ref. [83] provided some extended applications of facial curves. 3D curves were formed by intersecting three spheres with the 3D surface and used to compute adjustable integral kernels (RAIKs) in Ref. [83]. A sequence of RAIKs generated from the surface patch around each keypoint can be represented by 2D images such that certain characteristics of the represented 2D images have a positive impact on matching accuracy, speed, and robustness.

Ref. [88] introduced nasal patches and curves. First, seven landmarks in the nasal region were detected. A set of planes was created using pairs of landmarks. A set of spherical patches and curves were yielded by the intersection of these planes with the nasal surface to create the feature descriptor. Then the feature vectors were taken by concatenating histograms of x, y, and z components of the surface normal vectors of Gabor-wavelet filtered depth maps. Features were selected by a genetic algorithm for stability under changes in facial expression. Compared to previous methods, this method shows excellent separability. Recently, Ref. [93] presented a geometry and local shape descriptor based on the wave kernel signature (WKS) [103], to overcome distortions caused by facial expressions.

4.2.4 Region-based methods

A representative local descriptor is the local binary pattern (LBP) [104]. It was initially used for 2D images. Local geometric features extracted from certain regions of the face surface can be robust to face expression variations [13]. LBPs were used to represent the facial depth and normal information of each face region in Ref. [72], where a feature-based 3D face division pattern (see Fig. 11 in Ref. [72]) was proposed to reduce the influence of local facial distortion.

Recently, Ref. [95] used the LBP algorithm to extract features of 3D depth images and the SVM algorithm to classify them. The feature extraction time of each depth map in Texas-3D was reduced to 0.19 s while Ref. [70] required 23.54 s. Inspired by LBP, Ref. [76] proposed the multi-scale and multi-component local normal patterns (MSMC-LNP) descriptor, which can describe normal facial information more compactly. The Mesh-LBP method was used in Ref. [89], where LBP descriptors were directly computed on the 3D face mesh surface, fusing both shape and texture information.

Another type of local method is based on geometric features. Ref. [73] proposed a low-level geometric feature approach, which extracted region-based histogram descriptors from a facial scan. Feature regions include the nose and the eyes-and-forehead, which are comparatively less affected by the deformation caused by facial expressions. A support vector machine (SVM) and fusion of these descriptors at both feature and score level were applied to improve accuracy. In Ref. [79], a covariance matrix of features was used as the descriptor for 3D shape analysis, rather than the features themselves. Compared to feature-based vectors, covariance-based descriptors can fuse and encode all types of features into a compact representation [90]. Their work was expanded in Ref. [90].

There are other local methods. In Ref. [74], local surface descriptors were constructed around keypoints, which were defined by checking the curvelet coefficient in each sub-band. Each keypoint is represented by multiple attributes, such as curvelet position, direction, spatial position, scale, and size. A set of rotation-invariant local features can be obtained by rearranging the descriptors according to the orientation of the key points. The method in Ref. [84] used the regional boundary sphere descriptor (RBSR) to reduce the computational cost and improve the classification accuracy.

Ref. [91] proposed a local derivative pattern (LDP) descriptor based on local derivative changes. It can capture more detailed information than LBP. Recently, Yu et al. [105] recommended utilizing the iterative closest point (ICP) with resampling and denoising (RDICP) method to register each face patch

to achieve high registration accuracy. With rigid registration, all face patches can be used to recognize the face, significantly improving accuracy as they are less sensitive to expression or occlusion.

4.2.5 Summary

Most local methods can better handle facial expression and occlusion changes as they use salient points and rigid feature regions, such as nose and eyes, to recognize a face. The main objective of local methods is to extract distinctive compact features [13]. We summarize local methods as follows:

- Keypoint-based methods can process partial face images with missing parts or occlusion, since the feature representations are generated from a set of keypoints and their geometric relationships. However, if the number of keypoints is excessive, the computational cost increases; if the keypoints are too few, some key features will be lost and recognition performance is affected. In addition, algorithms for measuring the neighborhoods of keypoints play an important role as the geometric relationships of keypoints are used to build feature vectors.
- Most curved-based methods use radial curves since level curves are sensitive to occlusion and missing parts. Generally, a reference point is required in a curve-based method. The nose region is rigid and has more distinctive shape features than other regions, so the nose tip is used as the reference point in most curve-based methods [13]. Therefore, its detection is a crucial step. Inaccurate positioning of the nose tip can affect the extraction of curves and compromise results of the face recognition system.
- Most region-based methods are robust to changes in facial expression and pose as the feature vectors are extracted from rigid regions of the face surface. Some also need highly accurate nose tip detection as the nose tip is used for face segmentation.

4.3 Global methods

Unlike local methods, global methods extract features from the entire 3D face surface. They are very effective and can perform well given complete, frontal, fixed-expression 3D faces. Table 3 summarizes noteworthy endeavors in this area.

An intrinsic coordinate system for 3D face registration was proposed by Ref. [106]. This system is based on a vertical symmetry plane determined by the nose tip and nose orientation. A 3D point cloud surface is transformed into a face coordinate system and PCA-LDA is used to extract features from the range image obtained from the newly transformed data. Ref. [107] presented a method named UR3D-C, which used LDA to train the dataset and compress the biometric signature to only 57 coefficients. It still shows high discrimination with these compact feature vectors. Bounding sphere representation (BSR), introduced in Ref. [108], was used to represent both depth and 3D geometric shape information by projecting preprocessed 3D point clouds onto their bounding spheres.

Shape-based spherical harmonic features (SHF) were proposed in Ref. [109], where SHFs were calculated based on the spherical depth map (SDM). SHFs can capture the gross shape and fine surface details of a 3D face as the strengths of spherical harmonics at different frequencies. Ref. [110] used 2DPCA to extract features and employed Euclidean distances for matching. Ref. [111] proposed a computationally efficient and simple nose detection algorithm. It constructs a low-resolution wide-nose eigenface space using a set of training nose regions. A pixel in an input scan is determined to be the nose tip if the mean square error between the candidate feature vector and its projection on the Eigenface space is less than a predefined threshold.

Ref. [112] introduced a rigid-area orthogonal spectral regression (ROSR) method, where curvature information was used to segment rigid facial areas and OSR was utilized to extract discriminative features. In Ref. [113], a 3D point cloud was registered in the inherent coordinate system with the nose tip as the origin, and a two-layer ensemble classifier was used for face recognition. A local facial surface descriptor was proposed by Ref. [114]. This descriptor is constructed based on three principal curvatures estimated by asymptotic cones. The asymptotic cone is an essential extension of an asymptotic direction to a mesh model. It allows the generation of three principal curvatures representing the geometric characteristics of each vertex.

Ref. [115] proposed a region-based 3D deformable model (R3DM), which was formed from densely corresponding faces. Recently, kernel PCA was used for 3D face recognition. As faces exhibiting non-linear

Table 3 Global techniques

Author/year	Method	Advantage	Limitation	Database	RR1 (%)
Spreeuwers (2011) [106]	PCA-LDA	Less registration time	Expression, occlusion	FRGC v2	99
Ocegued et al. (2011) [107]	Ll norm, ICP, LDA, Simulated Annealing algorithm	Speed efficient	Expression, occlusion	FRGC v2	99.7
Ming and Ruan (2012) [108]	Robust group sparse regression model (RGSRM)	Expression, pose	Distorted images	FRGC v2 CASIA	— —
Liu et al. (2012) [109]	—	Faster, cost-effective	Expression, occlusion	SHREC2007 FRGC v2 Bosphorus	97.86 96.94 95.63
Taghizadegan et al. (2012) [110]	PCA, Euclidean distance	Expression	Occlusion	CASIA	98
Mohammadzade and Hatzinakos (2012) [111]	PCA	Computation, expression	Occlusion, pose	FRGC	—
Ming (2014) [112]	PCA, Spectral Regression, the orthogonal constraint	Expression, computational cost, storage space	Occlusion	FRGC v2	95.24
Ratyal et al. (2015) [113]	PCA, Mahalanobis Cosine (MahCos)	Pose, expression	Occlusion, missing part	GavabDB FRGC v2	100 98.93
Tang et al. (2015) [114]	Principal curvatures	Computational cost	Expression, occlusion	FRGC v2	93.16
Gilani et al. (2017) [115]	PCA, use CNN for landmark detection	Occlusion	Faster, expressions, poses	Bosphorus	98.1
Peter et al. (2019) [116]	Kernel-based PCA	Higher accuracy rate	—	FRGC v2	—

Table 4 Hybrid techniques

Author/year	Method	Advantage	Limitation	Database	RR1 (%)
Passalis et al. (2011) [117]	PCA	Pose, occlusion, missing data	Expression, low accuracy	UND	—
Huang et al. (2012) [118]	SIFT-based, extended LBP	Registration-free (frontal)	Large pose (alignment required)	FRGC v2	97.6
Alyüz et al. (2012) [119]	ICP, PCA, LDA	Occlusion	Expression	Bosphorus	83.99
Fadaifard et al. (2013) [120]	L1-norm	Noise, computational efficiency	Occlusion, expression	GavabDB	86.89
Alyüz et al. (2013) [121]	ICP, PCA, LDA	Occlusion, missing data	Expression	Bosphorus UMBDB	— —
Bagchi et al. (2014) [122]	ICP, PCA	Pose, occlusion	Pose	Bosphorus	91.3
Bagchi et al. (2015) [123]	ICP, KPCA	Pose	Expression	GavabDB Bosphorus FRAV3D	96.92 96.25 92.25
Liang et al. (2017) [124]	HK classification	Pose	Expression	Bosphorus	94.79

shapes, non-linear PCA was used in Ref. [116] to extract 3D face features, as it has notable benefits for data representation in high-dimensional space.

To sum up, most global methods have faster speed and lower computational demands, but they are unsuitable for handling occluded faces or faces with missing parts. In addition, variations in pose and scale may affect recognition accuracy when using global features, as global algorithms create discriminating features based on all visible facial shape information. This requires accurate normalization for pose and scale. However, it is not easy to perform accurate pose normalization given noisy or low-resolution 3D scanning.

4.4 Hybrid methods

Hybrid face recognition systems use both local features and global features. A comparison of recent hybrid methods is provided in Table 4.

Ref. [117] used an automatic landmark detector to estimate poses and detect occluded areas, and utilized facial symmetry to deal with missing data. Ref. [118] proposed a hybrid matching scheme using multiscale extended LBP and SIFT-based strategies. In Ref. [119], the problem of external occlusion was addressed and a two-step registration framework was proposed. First, a non-occluded model is selected for each face with the occluded parts removed. Then a set of non-occluded distinct regions is used to compute the masked projection. This method relies on accurate nose tip detection, and results are adversely affected if the nose area is occluded. Ref. [121] extended this work in 2013. Ref. [120] proposed a scale-space-based representation for 3D shape matching which is stable in the presence of surface noise.

In Ref. [122], Bagchi et al. used ICP to register a 3D range image and PCA to restore the occluded region. This method is robust to the noise and occlusion. Later, they improved the registration method and proposed an across-pose method in Ref. [123]. Ref. [124] also proposed a 3D face recognition method with pose-invariant and a coarse-to-fine approach to detect landmarks under large yaw variations. At the coarse search step, HK curvature analysis is used to detect candidate landmarks and subdivide them according to the classification strategy based on facial geometry. At the fine search step, the candidate landmarks are identified and marked by comparison with the face landmark model.

Hybrid 3D face recognition methods may use more complex structures than local or global methods, and as a result, may achieve better recognition accuracy at a higher computational cost. As in global methods, face registration is an important step for hybrid 3D methods, especially for overcoming pose variation and occlusion.

5 Deep learning-based 3D face recognition

5.1 Overview

In the last decade, deep neural networks have become one of the most popular approaches for face recognition. Compared to conventional approaches, deep learning-based methods have great advantages over image processing [125]. For conventional methods, the key step is to find robust feature points and descriptors based on geometric information in 3D face data [51]. Compared to end-to-end deep learning models, these methods have good recognition performance, but involve relatively complex algorithmic operations to detect key features [51]. For deep learning-based methods, robust face representations can be learned by training a deep neural network on large datasets [51], which can hugely improve face recognition speed.

There are a variety of deep neural networks for facial recognition. Convolutional neural networks (CNN) are the most popular. The robust and discriminative feature representations learned via CNN can significantly improve the accuracy of face recognition, as demonstrated by Refs. [42, 51]. Recently, graph convolutional networks (GCNs) have also been considered in the face recognition field to solve the problem of large face deformation in real life. GCNs utilize filters to identify high-level similarities between nodes by extracting high-dimensional features of nodes and their neighborhoods in a graph [58].

Figure 4 depicts a common face recognition process based on a deep-CNN (DCNN). In the training phase, the training dataset is preprocessed (see Section 3) to generate a unified feature map. The feature map is resized to fit the input tensor to the DCNN architecture (in terms of the height, width, and number of channels of the input network layer, and the number of images). Then, the DCNN is trained with the preprocessed maps. In the testing phase, a 3D face scan is selected for each identity from the test dataset as the identity dataset. The feature representations of the identity dataset are obtained through the trained network as the feature library. Then, the feature vector of the probes is obtained from the trained DCNN and used to match the features in the given gallery. In the matching process, the feature gallery is scanned and the distance between each feature representation and the feature vector of the probed surface is calculated. The identity with the closest matching distance is then returned.

With the application of CNN, the accuracy of 2D

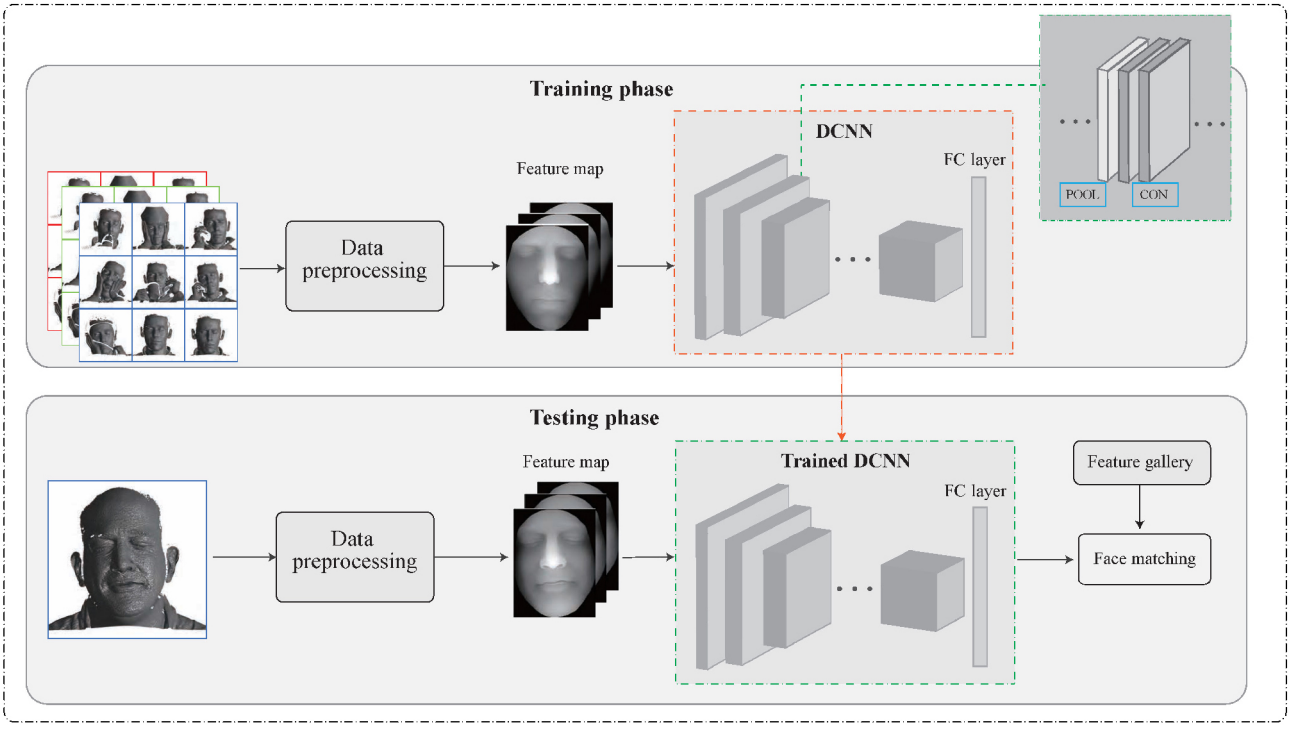


Fig. 4 Overview of 3D deep learning-based face recognition methods.

face recognition systems (DeepFace [126], DeepID series [127–130], VGG-Face [16], FaceNet [131]) has significantly improved. In these systems, face representations are learned directly from 2D facial images by training deep neural networks on large datasets. Accuracy is close to 100% on some specific databases (such as LFW). The high recognition rate for 2D face recognition shows that CNN-based methods are superior to the conventional feature extraction methods. Based on the intrinsic advantages of 3D faces relative to 2D faces in handling uncontrolled conditions such as changes in pose, illumination, and expression, researchers are attracted to applying DCNNs to 3D face recognition.

Indeed, some of these 2D face recognition networks are still being used by some 3D methods. In such 3D face recognition methods, 3D faces are converted into 2D maps as input of the network. Other networks directly accept 3D data as input, such as PointNet [132], PointNet++ [133]. Based on the input formats of the network, we classify deep learning-based 3D face recognition methods into three categories: 2D-input, 3D-input, and graph-input networks. Table 5 lists these approaches; details of these methods are discussed below.

Table 5 Network architectures based on 3D deep learning techniques

Category	Backbone/architecture	Reference
2D-input	VGG-Face	[51, 134]
	ResNet	[5, 135, 136]
	MobileNet	[137]
	Others	[42, 138–144]
3D-input	PointNet++, PointFace	[145–148]
Graph-input	GCN	[149]

5.2 2D-input networks

Kim et al. [51] proposed the first 3D face recognition model with DCNN. They adopted VGG-Face [16] pre-trained on 2D face images as their network, and then fine-tuned the network with augmented 2D depth maps. The last FC layer of VGG-Face is replaced by a new last FC layer and a softmax layer. In the new last layer, weights are randomly initialized using a Gaussian distribution with a mean of zero and a standard deviation of 0.01. The size of the dataset is expanded by augmenting the 3D point cloud of face scans with expression and pose variations during the training phase. A multi-linear 3DMM is used to generate more data, including variations in both

Table 6 Deep learning-based methods

Author/year	Network	Data augmentation	Loss function	Matching	Database	RR1 (%)
Kim et al. (2017) [51]	Fine-tuning VGG-Face	3DMM	—	Cosine distance	Bosphorus	99.2
Zulqarnain Gilani and Mian (2018) [42]	FR3DNet	Synthesized from original 3D face pairs	—	Cosine distance	Texas-3D	100.0
Ding et al. (2019) [134]	Fine-tuning VGG-Face	—	—	SVM	CurtinFace [150]	93.41
Feng et al. (2019) [139]	ANN	—	—	—	CASIA	98.44
Xu et al. (2019) [138]	LeNet5	—	—	Euclidean distance	CASIA	—
Tan et al. (2019) [136]	ResNet-18	Randomly occlude depth-maps with 1-6 patches	AMSoftmax [151]	Cosine distance	CASIA	99.7
Mu et al. (2019) [140]	MSFF	Pose augmentation, shape jittering, shape scaling	Cross entropy loss	Cosine distance	Lock3DFace	84.22
Lin et al. (2019) [135]	ResNet-18	Feature tensor based augmentation	Binary cross-entropy loss	Similarity tensor calculated from 2 feature tensors	Bosphorus	99.71
Olivetti et al. (2020) [137]	MobileNetV2	Rotate the original depth map	—	—	Bosphorus	97.56
Dutta et al. (2020) [141]	SpPCANet	—	—	Linear SVM [152]	Frav3D	96.93
Lin et al. (2021) [143]	MQFNet	Generate depth image by pix2pix [153]	Weighted loss	—	Lock3DFace	86.55
Cai et al. (2019) [5]	Pre-ResNet-34, Pre-ResNet-24, Pre-ResNet-14	Resolution augmentation	Multi-scale triplet loss	Euclidean distance	FRGC v2	100
Cao et al. (2020) [142]	ANN	—	—	—	H3D [142]	—
Chiu et al. (2021) [144]	Mask-guided RGB-D Net	Generate depth image by DepthNet	Attribute-aware loss [154]	Cosine distance	Lock3DFace	96.43
Bhople and Prakash (2021) [147]	Triplet Net	P-level data augmentation	Triplet loss	Squared distances	Bosphorus	97.55
Jiang et al. (2021) [148]	PointFace	Random crop	Feature similarity loss	Cosine distance	Lock3DFace	87.18
Zhang et al. (2021) [145]	PointNet++	GPMM-based	Triplet loss	Cosine similarity	Bosphorus	99.68
Papadopoulos et al. (2021) [149]	Face-GCN	—	—	—	BU4DFE	88.45

shape (α) and expression (β). A 3D point cloud can be represented by

$$\boldsymbol{X} = \overline{\boldsymbol{X}} + \boldsymbol{P}_{\mathrm{s}}\alpha + \boldsymbol{P}_{\mathrm{e}}\beta \tag{1}$$

where $\overline{\boldsymbol{X}}$ is the average facial point cloud, $\boldsymbol{P}_{\mathrm{s}}$ is the shape information provided by the Basel Face Model [155], and $\boldsymbol{P}_{\mathrm{e}}$ is an expression provided by FaceWarehouse [156]. Expression variations are created by randomly changing the expression β parameter in the 3DMM. Randomly generated rigid transformations are applied to the input 3D point cloud to produce pose variations. During data preprocessing, a nose tip is first found in the 3D point cloud, and then the 3D point cloud is cropped within a 100 mm radius. Classical rigid-ICP [157] between the cropped 3D data and the reference face model is used to align the 3D data. In order to fit the input size of the CNN architecture, the aligned 3D data are orthogonally projected to 2D images to generate a $224\times224\times3$ depth map. In addition, eight 18×18 patches are randomly placed on the depth

map to simulate occlusion and prevent overfitting to specific regions of the face. The model was evaluated on three public 3D databases: Bosphorus [30], BU3D-FE [26], and 3D-TEC [35], yielding recognition rates of 99.2%, 95.0%, and 94.8%, respectively.

A deep 3D face recognition network (FR3DNet) [42] was trained on 3.1 million 3D faces; it is specifically designed for 3D face recognition. It is also based on VGG-Face [16]. A rectifier layer is added for every convolutional layer. Compared to Kim et al.'s work [51], a much larger dataset is generated and expanded by new identities. A new face $\hat{\boldsymbol{F}}$ is generated from a pair of faces ($\boldsymbol{F}_i$, $\boldsymbol{F}_j$) with the maximum non-rigid shape difference:

$$\hat{\boldsymbol{F}} = (\boldsymbol{F}_i + \boldsymbol{F}_j)/2 \tag{2}$$

The synthetic faces generated by this method have richer shape changes and details than statistical face models [155]. However, the computational cost is very high as they are all generated from high-dimensional raw 3D faces. In addition, 15 synthetic cameras are deployed in the frontal hemisphere of the 3D face to simulate pose variations and occlusions in each 3D scan. To fit the input of FR3DNet, the 3D point cloud data are preprocessed to a $160 \times 160 \times 3$ image [155]. Before aligning and cropping the face, the point cloud is converted into a three-channel image. These three channels indicate three surfaces generated by using the gridfit algorithm [158] to give depth map $z(x, y)$, azimuth map $\theta(x, y)$, and elevation map $\phi(x, y)$, where θ and ϕ are azimuth and elevation angles of the normal vectors of the 3D point cloud surface, respectively. Experiments were conducted on most public databases; the highest recognition accuracy was achieved on the Texas-3D database, reaching 100%.

Ding et al. [134] proposed an SVM based RGB-D face recognition algorithm combining 2D color and 3D depth features. A fine-tuned VGG-face network is used to determine 2D features from color images. 3D geometric features are obtained by computing expression-invariant geodesic distances between facial landmarks on a 3D face mesh. The 2D and 3D features are then used as RGB-D classifiers to train the SVM. Experiments were performed on the CurtinFace [150] database, achieving good results with pose variations (93.41%) and neutral expression variations (100%).

Feng et al. [139] adopted a two-DCNN module to extract features from color images and depth maps built from 3D raw data. The outputs of the two feature layers are fused as the final input to an artificial neural network (ANN) recognition system. It was tested on CASIA (V1) and compared recognition rates using the 2D feature layer, 3D feature layer, and the fusion of both layers. A higher RR1 (98.44%) was obtained with fused features.

Xu et al. [138] also designed a dual neural network to reduce the number of training samples needed. The network consists of a dual-channel input layer that can fuse a 2D texture image and a 3D depth map into one channel, and two parallel LeNet5-based CNNs. Each CNN processes the fused image separately to obtain its feature maps, which are used to calculate similarity. The gray-scale depth map obtained from the point cloud, combined with the corresponding 2D texture, is used as the dual-channel input. The most important preprocessing step is face hole filling, to provide a more intact face. The basic idea is to first extract the 3D hole edge points, then project the hole edge points onto the 2D mesh plane to fill the hole points, and map them back to the original 3D point cloud. Experiments were conducted to show the influence of depth map features and the size of training set on the accuracy of recognition rate.

Tan et al. [136] designed a framework to specifically process the low-quality 3D data captured by portable 3D acquisition hardware such as mobile phones. The framework includes two parts: face registration and face recognition. At the face registration stage, a PointNet-like deep registration network (DRNet) is used to reconstruct the dense 3D point cloud from low-quality sequences. The DRNet is based on ResNet-18 and takes a pair of $256 \times 256 \times 3$ coordinate-maps as input. To obtain the desired sparse samples from the raw datasets, noise and random pose variation are added to the face scan. Then the new point cloud is projected onto a 2D plane with 1000 grid cells of the same size. A sparse face of 1000 points is obtained by randomly selecting a point from each cell. Six sparse faces are generated from each face scan and passed to DRNet to generate a new dense point cloud. Then the fused data are used as the input to a face recognition network (FRNet) also based on ResNet-18. Compared to FR3DNet, its facial RR1 on UMBDB is higher, reaching 99.2%.

Mu et al. [140] proposed a lightweight CNN for 3D face recognition, especially for low-quality data.

This network contains 4 blocks with 32, 64, 128, and 256 convolution filters. The feature maps from these four convolutional blocks are captured by different receptive fields, downsampled to fixed size by max-pooling and integrated to form another conversion block. This process is completed by a multi-scale feature fusion module. The aim is to efficiently improve the representation of low-quality face data. A spatial attention vectorization (SAV) module is used to replace the global average pooling layer (also used by ResNet) to vectorize feature maps. The SAV highlights important spatial facial clues and conveys more discriminative cues by adding an attention weight map to each feature map. In addition, three methods are used to augment the training data: pose generation (by adjusting virtual camera parameters), shape jittering (by adding Gaussian noise to simulate rough surface changes), and shape scaling (by zooming in 1.1× to the depth face image). During data preprocessing, as in the above methods, a 10×10 patch surface is first cropped around the given nose tip with outlier removal. Then the cropped 3D point cloud is projected onto a 2D space (depth surface) to generate a normal map image.

Lin et al. [143] also designed a multi-quality fusion network (MQFNet) to improve the performance of low-quality 3D face recognition. First, the pix2pix network [153] is used to generate high-quality depth maps from low-quality faces. To avoid the effect of loss of identity features in images generated by pix2pix, MQFNet contains two pipelines to extract and fuse features from low-quality and high-quality images and to generate more discriminative features. This work was also tested on the Lock3DFace database. Compared to Ref. [140], the average accuracy was improved by 8.11%.

Olivetti et al. [137] proposed a method based on MobileNetV2. MobileNet is a comparatively new neural network specifically designed for mobile phones. It is easy to train and requires only a few parameters to be tuned. This work was based on the Bosphorus database, which only contains 105 identities with 4666 images. To obtain sufficient training samples, they augmented the data by rotating the original depth map (clockwise 25^{circ}, counterclockwise 40^{circ}) and horizontally minoring each depth map. The most important part of their work is the input data for DCNN. Geometric descriptors are used as input data instead of pure facial depth maps. The selection of geometric feature descriptors is based on the GH-EXIN network. The reliability of geometric descriptors based on curvature is demonstrated in Ref. [159]. The input is a three-channel image including the 3D facial depth map, the shape index, and the curvedness, which enhances the accuracy of the network. A 97.56% RR1 was achieved on the Bosphorus database.

Dutta et al. [141] also proposed a lightweight sparse principal component analysis network (SpPCANet). It includes three parts: a convolutional layer, a nonlinear processing layer, and a feature merging layer. For data preprocessing, usual methods are used to detect and crop the face area. First, an ICP-based registration technology is used to register a 3D point cloud, and then the 3D point cloud is converted into a depth image. Finally all faces are cropped to rectangles based on the position of the nose tip. The system obtained a 98.54% RR1 on Bosphorus.

Lin et al. [135] adopted ResNet-18 [17] as the backbone of their network. The big difference from other work is their data augmentation method. Instead of generating 3D face samples from 2D face images, they generated feature tensors directly based on Voronoi diagram subdivision. The salient points are detected from a 3D face point cloud with its corresponding 2D face image and divided into 13 subdivisions based on the Voronoi diagram. The face can be expressed as $\boldsymbol{F} = [f_1, \cdots, f_{13}]$ and the sub-feature is $\mathbf{Sub}\boldsymbol{F}_i$. The feature tensor is extracted from a 3D mesh by detecting the salient points and integrating features of all the salient points, which can be represented as

$$\boldsymbol{F}^k = \cup_{i=1}^{13} \mathbf{Sub}\boldsymbol{F}_i^k, \quad k = 1, \cdots, K \tag{3}$$

where K is the number of 3D face samples of the same person. A new feature set can be synthesized by randomly choosing the ith sub-feature set from the K samples. Excellent results were achieved on both Bosphorus and BU3D-FE databases with accuracies of 99.71% and 96.2%, respectively.

Cai et al. [5] designed three deep residual networks with different layers based on ResNet: Pre-ResNet-14, Pre-ResNet-24, and Pre-ResNet-34. Multi-scale triplet loss supervision is constructed by combining a softmax loss and the two triplet losses as supervision on the last fully connected layer and the last feature layer. To enlarge the size of the training set, the data

are augmented in three ways: pose augmentation based on 3D scans, resolution augmentation, and transformational augmentation based on the range images. For the preprocessing algorithm, raw 3D data are converted into a 96×96 range image and only the center of the two pupils and nose tip are used for alignment. For the preprocessing algorithm, three overlapping face components (the upper half face, the small upper half face, and the nose tip) and the entire facial region are generated from the raw 3D data. The most important part of this method is detecting the nose tip and two pupils. Three landmarks are detected from the 2D textured image of the corresponding 3D face data and are mapped to the 3D model. Then, a new nose tip is calculated by taking the highest point of the nose region (centered on the tip of the nose with a radius of 25 mm). The nose tip is re-detected on the 3D model as the 2D domain detection may reduce detection accuracy due to pose variations. Another reason for detecting the nose tip by this means is that the lower dimensional feature vectors generated can be used to detect the new nose tip to reduce computational cost. Finally, the feature vectors of the four patches can be used alone or in combination for matching. It obtained high accuracy on four public 3D face databases: FRGC v2, Bosphorus, BU-3DFE, and 3D-TEC, with 100%, 99.75%, 99.88%, and 99.07%, respectively.

Cao et al. [142] believed that the key to a reliable face recognition system is rich data sources and paid more attention to data acquisition. Therefore, a holoscopic 3D (H3D) face image database was created, which contains 154 raw H3D images. H3D imaging was recorded by using a regularly closely packed array of small lenses connected to a recording device. It can display 3D images with continuous parallax and full-color images can be viewed from a wider viewing area. A wavelet transform is used for feature extraction, as it performs well in the presence of illumination change and face orientation change, reducing image information redundancy and retaining the most important facial features. While this is a new direction for 3D face recognition, the accuracy of this method is quite low, only reaching just over 80% on the H3D database.

Chiu et al. [144] applied an attention mechanism in a face recognition system that has two main parts: a depth estimation module (DepthNet) and a mask-guided face recognition module. DepthNet is used to convert 2D face datasets to RGB-D face datasets and to address the a lack of identities in 3D face datasets for network training. The augmented 3D database is used to train a mask-guided RGB-D face recognition network, which has a two-stream multi-head architecture with three branches: an RGB recognition branch, a depth map recognition branch, and an auxiliary segmentation mask branch with a spatial attention module. The latter shares weights between the two recognition branches and provides auxiliary information from the segmentation branch to help the recognition branches focus on informative parts such as eyes, nose, eyebrows, and lips. This module achieved good results on multiple databases and had a higher average accuracy (96.43%) on Lock3DFace than Refs. [140, 143].

5.3 3D-input networks

Bhople et al. [146] utilized a Siamese network with PointNet-CNN (PointNet implementation with CNN) to determine the similarity and dissimilarity of point cloud data. In Ref. [147], they continued their work and proposed a triplet network with triplet loss, in a variant of the Siamese network. The triplet network is a shared network consisting of three parallel symmetric CPNs using the PointNet architecture and CNN. The input to the network is a triplet of three 3D face scans: positive, anchor, and negative. Since the input is in the form of a triplet, more information is captured during training. The Siamese network and the triplet network were compared in terms of recognition rates on the Bosphorus and IIT Indore databases. The triplet network achieved better accuracy on the IIT Indore database, but not on Bosphorus. They also performed point cloud-level data augmentation by rotating the point cloud data in a fixed orientation by randomly perturbing the points by a small rotation and jittering the position of each point slightly.

PointFace [148] consists of two weight-shared encoders that extract discriminative features from a pair of point cloud faces. In the training phase, each encoder learns identity information from each sample itself, while using feature similarity loss to evaluate the embedding similarity of two samples. The feature similarity loss function can be represented by

$$L_{\text{sim}} = \sum_{i=1}^{M} [D(f_i^{\text{a}}, f_i^{\text{p}}) + m - D(f_i^{\text{a}}, f_i^{\text{n}})] \quad (4)$$

where $f_i^{\mathrm{a}}, f_i^{\mathrm{p}}$, and f_i^{n}, $i = 1, \cdots, M$, are the L_2 normalized feature vectors of the anchor, positive, and negative samples, respectively. $D(.,.)$ is the distance between two vectors. The encoder can distinguish 3D faces from different individuals and compactly cluster features coming from faces of the same person. Compared to Refs. [140, 143], the model has higher average accuracy (87.18%) on Lock3DFace.

Zhang et al. [145] proposed a 3D face point cloud recognition framework based on PointNet++. It consists of three modules: training data generator, face point cloud network, and transfer learning. The most important part of this work is the training data generator. All training sets are unreal data, synthesized by sampling from a statistical 3DMM of face shape and expression based on a GPMM [160]. This method addresses the problem of lack of a large training dataset. After classification training, triplet loss is used to fine-tune the network with real faces to give better results.

5.4 Graph-input networks

Papadopoulos et al. [149] introduced a registration-free method for dynamic 3D face recognition based on spatiotemporal graph convolutional networks (ST-GCN). First, facial landmarks are estimated from a 3D mesh. Landmarks alone are insufficient for facial recognition because crucial geometric and texture information are left out. To describe local facial patterns around landmarks, new points are first interpolated between the estimated landmarks, and then a kD-tree search is used to find the closest points to each landmark. For each frame, a facial landmark corresponds to a vertex ($\boldsymbol{v}_i \in V$) in a graph ($\boldsymbol{G} = V, E$), and the landmarks are connected as spatial edges according to a defined relationship. For the 3D sequences of meshes, identical landmarks are connected in consecutive frames as temporal edges. This work was tested on BU4DFE with an average accuracy of 88.45%. The performance is not as good as other state-of-the-art methods [42, 161, 162], but it demonstrates the feasibility of using GCN for dynamic 3D face recognition.

5.5 Summary

This section has reviewed deep learning-based 3D face recognition techniques and classified them into three categories based on their network input formats. Most deep learning-based 3D methods achieve high recognition accuracy and run quickly. For example, Ref. [42] gets 100% RR1 on Texas-3D and Ref. [5] only requires 0.84 s to identify a target face from a gallery of 466 faces. There are three important parts in a deep learning-based system: data preprocessing, data augmentation, and network architecture. Usually, the input data need to be preprocessed (face registration) to find correspondences between all vertices of the face mesh, since CNNs are usually intolerant of pose changes. Deep learning-based methods always require a large amount of data to train the network, especially if training the network from scratch. To avoid this, some works [51, 163] transfer learning from a pre-trained model and fine-tune the network on a small dataset, which also takes less training time. But lacking large-scale 3D face datasets is still an open problem for DCNN-based 3D face recognition research. Data augmentation is an important method to enlarge 3D face databases by generating new faces from existing ones. In addition, adopting a suitable network is important. Most of the above-reviewed works use a single CNN but a few use dual CNNs, such as Ref. [138]. As more networks are adopted in this field, reorganization of existing network architectures may also be a topic of future research.

6 Discussion

In the past decade, 3D face recognition has achieved significant advances in 3D face databases, recognition rates, and robustness to face data variations, such as low-resolution, expression, pose, and occlusion. In this paper, conventional methods and deep learning-based methods have been thoroughly reviewed in Sections 4 and 5, respectively. Based on the feature extraction algorithms, conventional methods are divided into three types: local, global, and hybrid methods.

- Local feature descriptors extract features from small regions of a 3D facial surface. In some cases, the region can be reduced to small patches around detected keypoints. The number of extracted local descriptors is related to the content of the input face (entire or partial). It is commonly assumed that only a small number of facial regions are affected by occlusion, missing data, or distortion caused by data corruption, while most other regions persist unchanged. Face representation is derived from a combination of

many local descriptors. Therefore, local facial descriptors are not compromised when dealing with changes to a few parts caused by facial expressions or occlusion [87].

- A global representation is extracted from an entire 3D face, which usually makes global methods compact and therefore computationally efficient. While these methods can achieve great accuracy in the presence of complete neutral faces, they rely on the availability of full face scans and are sensitive to face alignment, occlusion, and data corruption.
- Hybrid methods can handle more conditions, such as pose and occlusion variations.

Since 2016, much research on deep learning-based 3D face recognition has been carried out. Table 7 summarizes the RR1 of our surveyed methods tested on different databases. Compared to conventional face recognition algorithms, deep learning-based methods have the advantages of simpler pipelines and greater accuracy.

To improve the accuracy and performance of face recognition systems, the following (future) directions are suggested, concerning new face data generation, data preprocessing, network design, and loss functions.

- *Large-scale 3D face databases.* Current 3D face databases are often smaller than their counterparts in 2D color face recognition; nearly all the deep learning-based 3D face recognition methods fine-tune pre-trained networks on converted data from 3D faces. Larger-scale 3D face databases could enable training from scratch and improve recognition difficulty, closing the gap to real-world applications.
- *Augmenting face data.* As Section 5 notes, almost every proposed method provides a strategy for augmenting face training data, as a large amount of training data are required to train networks. A network trained with sufficient data can better distinguish features, while a small number of samples may result in overfitting. We can increase the size of the 3D database by generating more images for existing identities or synthesizing new

Table 7 RR1 (%) of deep learning-based methods on various databases. H=high-quality image. L=low-quality image. F=fine-tuning

Reference	FRGC v2	BU3D-FE	BU4D-FE	Bosphorus	CASIA	GavabDB	Texas-3D	3D-TEC	UMBDB	ND-2006	Lock3DFace
[51]	—	95.00	—	99.20	—	—	—	94.80	—	—	—
[42]	97.06	98.64	95.53	96.18	98.37	96.39	100	97.90	91.17	95.62	—
[42] (F)	99.88	99.96	98.04	100	99.74	99.70	100	99.12	97.20	99.13	—
[139]	—	—	—	—	85.93	—	—	—	—	—	—
[5]	100	99.88	—	99.75	—	—	—	99.07	—	—	—
[135]	—	96.20	—	99.71	—	—	—	—	—	—	—
[137]	—	—	—	97.56	—	—	—	—	—	—	—
[136]	—	—	—	99.20	99.70	—	—	—	99.2	—	—
[52]	92.74	—	—	93.38	—	—	—	—	—	—	—
[52] (F)	98.73	—	—	97.50	—	—	—	—	—	—	—
[140] (H)	—	—	—	91.27	—	—	—	—	—	—	—
[140] (L)	—	—	—	90.70	—	—	—	—	—	—	—
[141]	—	—	—	98.54	88.80	—	—	—	—	—	—
[143]	—	—	—	—	—	—	—	—	—	—	86.55
[144]	99.27	100	—	100	100	—	—	—	—	—	96.43
[147]	—	—	—	97.55	—	—	—	—	—	—	—
[145]	99.60	—	—	99.68	—	—	—	—	—	—	—
[148]	—	—	—	—	—	—	—	—	—	—	87.18
[149]	—	—	88.45	—	—	—	—	—	—	—	—

identities. Common ways to generate new images are: rotating and cropping existing 3D data, or using 3DMM to slightly change the expression. To generate new identities, some model is designed to synthesize new faces from existing identities [42, 155, 164, 165]. Recently, generative adversarial networks (GANs) have been used for face augmentation where a face simulator is trained to generate realistic synthetic images. Recent works are summarized in Section 3.

- *Data preprocessing.* This is also key to improving face recognition accuracy. Besides removing redundant information, another goal of data preprocessing is to perform registration. A well-known problem of rigid-ICP registration is that it cannot guarantee optimal convergence [51]: it may not be possible to accurately register all 3D faces in different poses to the reference face. For 3D input networks, all 3D faces are taken as point-to-point correspondences for non-rigid registration. As an alternative, Ref. [149] proposed a registration-free method based on GCN to avoid this step. However, further work is needed to improve face recognition performance.
- *Data conversion.* As Section 5 explains, some works are based on 2D-input networks. To use them, better conversion techniques (e.g., from 3D faces to 2D maps) would improve face recognition performance.
- *Network architecture.* Many networks are available for 3D face recognition (see Table 5). Some researchers directly adopt pre-trained networks and then fine-tune them using training data generated from 3D faces, which can greatly improve the training speed. Also, dual or multiple networks can be used to handle different tasks, as did Refs. [5, 138].
- *Appropriate loss functions.* Using effective loss functions can reduce the complexity of training and improve feature learning capabilities. Most loss functions share a similar basic idea and aim to facilitate the training process by amplifying discriminative features from different individuals and compacting clustering features from the same individual. A commonly used loss function is the softmax loss, which encourages separability between classes but is incapable of supporting compactness within classes. For face recognition, highly discriminative features are required because the difference between two faces may be small, such as in twins. Therefore, applying loss functions to supervise the network layers has become one active research topic. For example, Ref. [5] adopted multi-scale loss supervision to improve extraction efficiency by combining one softmax loss and two triplet losses.

In addition to the above issues, researchers can consider combining conventional methods with CNNs. For example, keypoint detection techniques in conventional 3D face recognition methods could be incorporated into the deep learning-based methods to better pay attention to the area of interest. 3D face recognition methods for low-quality (low-resolution) data also need more work.

To apply 3D face technology to real-world applications, several things need to be considered: recognition time, quality of the input data, and pose and expression variations of the subject. Lightweight networks [140, 141] can reduce recognition time and improve efficiency. In Refs. [136, 140, 143], representations of low-quality face data are improved by fusing features from high-quality images. To handle pose and expression variations, the network can be trained using face datasets with rich expressions and pose changes to improve its robustness. Furthermore, dynamic 3D face recognition using 3D face sequences as input should be considered in the future.

7 Conclusions

3D face recognition has become an active and popular research topic in the field of image processing and computer vision in recent years. In this paper, a summary of public 3D face databases is first provided, followed by a comprehensive survey on 3D face recognition methods proposed in the past decade. They are divided into two categories based on their feature extraction methods: conventional and deep learning-based.

Conventional techniques are further classified into local, global, and hybrid methods. We have reviewed these methods by comparing their performance on different databases, computational cost, and robustness to expression change, occlusion, and pose variation. Local methods can better handle face expressions and occluded images at the cost of greater

computation than global methods. Hybrid methods can achieve better results and address challenges such as pose variation, illumination change, and facial expressions.

We have reviewed recent advances in 3D face recognition based on deep learning, mainly focusing on face augmentation, data preprocessing, network architecture, and loss functions. According to the input formats of the network adopted, the deep learning-based 3D face recognition methods may be broadly divided into 2D-input, 3D-input, and graph-input networks. With these powerful networks, the performance of 3D face recognition has been greatly improved.

We have also discussed the characteristics and challenges involved, and have provided potential future directions for 3D face recognition. For instance, large-scale 3D face databases are greatly needed to advance 3D face recognition in the future. We believe our survey will provide valuable information and insight to readers and the community.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Patil, H.; Kothari, A.; Bhurchandi, K. 3-D face recognition: Features, databases, algorithms and challenges. *Artificial Intelligence Review* Vol. 44, No. 3, 393–441, 2015.

[2] Zhou, H. L.; Mian, A.; Wei, L.; Creighton, D.; Hossny, M.; Nahavandi, S. Recent advances on singlemodal and multimodal face recognition: A survey. *IEEE Transactions on Human-Machine Systems* Vol. 44, No. 6, 701–716, 2014.

[3] Bowyer, K. W.; Chang, K.; Flynn, P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding* Vol. 101, No. 1, 1–15, 2006.

[4] Huang, G. B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Proceedings of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.

[5] Cai, Y.; Lei, Y. J.; Yang, M. L.; You, Z. S.; Shan, S. G. A fast and robust 3D face recognition approach based on deeply learned face representation. *Neurocomputing* Vol. 363, 375–397, 2019.

[6] Zhou, S.; Xiao, S. 3D face recognition: A survey. *Human-Centric Computing and Information Sciences* Vol. 8, No. 1, 35, 2018.

[7] Blackburn, D. M.; Bone, M.; Phillips, P. J. Face recognition vendor test 2000: Evaluation report. Technical report. Defense Advanced Research Projects Agency Arlington VA, 2001.

[8] Phillips, P. J.; Grother, P.; Micheals, R.; Blackburn, D. M.; Tabassi, E.; Bone, M. Face recognition vendor test 2002. In: Proceedings of the IEEE International SOI Conference, 44, 2003.

[9] 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 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 947–954, 2005.

[10] Phillips, P. J.; Scruggs, W. T.; O'Toole, A. J.; Flynn, P. J.; Bowyer, K. W.; Schott, C. L.; Sharpe, M. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 32, No. 5, 831–846, 2010.

[11] Abate, A. F.; Nappi, M.; Riccio, D.; Sabatino, G. 2D and 3D face recognition: A survey. *Pattern Recognition Letters* Vol. 28, No. 14, 1885–1906, 2007.

[12] Smeets, D.; Claes, P.; Hermans, J.; Vandermeulen, D.; Suetens, P. A comparative study of 3-D face recognition under expression variations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* (*Applications and Reviews*) Vol. 42, No. 5, 710–727, 2012.

[13] Soltanpour, S.; Boufama, B.; Jonathan Wu, Q. M. A survey of local feature methods for 3D face recognition. *Pattern Recognition* Vol. 72, 391–406, 2017.

[14] Guo, G. D.; Zhang, N. A survey on deep learning based face recognition. *Computer Vision and Image Understanding* Vol. 189, 102805, 2019.

[15] Masi, I.; Wu, Y.; Hassner, T.; Natarajan, P. Deep face recognition: A survey. In: Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images, 471–478, 2018.

[16] Parkhi, O. M.; Vedaldi, A.; Zisserman, A. Deep face recognition. In: Proceedings of the British Machine Vision Conference, 1–12, 2015.

[17] He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778, 2016.

[18] Lawrence, S.; Giles, C. L.; Tsoi, A. C.; Back, A. D. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks* Vol. 8, No. 1, 98–113, 1997.

[19] Howard, A.; Zhmoginov, A.; Chen, L. C.; Sandler, M.; Zhu, M. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *arXiv preprint* arXiv:1801.04381, 2018.

[20] Colombo, A.; Cusano, C.; Schettini, R. UMB-DB: A database of partially occluded 3D faces. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2113–2119, 2011.

[21] Beumier, C.; Acheroy, M. Automatic 3D face authentication. *Image and Vision Computing* Vol. 18, No. 4, 315–321, 2000.

[22] Hesher, C.; Srivastava, A.; Erlebacher, G. A novel technique for face recognition using range imaging. In: Proceedings of the 7th International Symposium on Signal Processing and Its Applications, 201–204, 2003.

[23] Moreno, A. GavabDB: A 3D face database. In: Proceedings of the 2nd COST275 Workshop on Biometrics on the Internet, 75–80, 2004.

[24] Chang, K. I.; Bowyer, K. W.; Flynn, P. J. An evaluation of multimodal 2D 3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 27, No. 4, 619–624, 2005.

[25] Wang, Y. M.; Pan, G.; Wu, Z. H.; Wang, Y. G. Exploring facial expression effects in 3D face recognition using partial ICP. In: *Computer Vision – ACCV 2006. Lecture Notes in Computer Science, Vol. 3851*. Narayanan, P. J.; Nayar, S. K.; Shum, H. Y. Eds. Springer Berlin Heidelberg, 581–590, 2006.

[26] Yin, L. J.; Wei, X. Z.; Sun, Y.; Wang, J.; Rosato, M. J. A 3D facial expression database for facial behavior research. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 211–216, 2006.

[27] Xu, C. H.; Tan, T. N.; Li, S.; Wang, Y. H.; Zhong, C. Learning effective intrinsic features to boost 3D-based face recognition. In: *Computer Vision – ECCV 2006. Lecture Notes in Computer Science, Vol. 3952*. Leonardis, A.; Bischof, H.; Pinz, A. Eds. Springer Berlin Heidelberg, 416–427, 2006.

[28] Conde, C.; Serrano, A.; Cabello, E. Multimodal 2D, 2.5D & 3D face verification. In: Proceedings of the International Conference on Image Processing, 2061–2064, 2006.

[29] Faltemier, T. C.; Bowyer, K. W.; Flynn, P. J. Using a multi-instance enrollment representation to improve 3D face recognition. In: Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems, 1–6, 2007.

[30] Savran, A.; Alyüz, N.; Dibeklioğlu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. Bosphorus database for 3D face analysis. In: *Biometrics and Identity Management. Lecture Notes in Computer Science, Vol. 5372*. Schouten, B.; Juul, N. C.; Drygajlo, A.; Tistarelli, M. Eds. Springer Berlin Heidelberg, 47–56, 2008.

[31] Heseltine, T.; Pears, N.; Austin, J. Three-dimensional face recognition using combinations of surface feature map subspace components. *Image and Vision Computing* Vol. 26, No. 3, 382–396, 2008.

[32] Ter Haar, F. B.; Daoudi, M.; Veltkamp, R. C. SHape REtrieval contest 2008: 3D face scans. In: Proceedings of the IEEE International Conference on Shape Modeling and Applications, 225–226, 2008.

[33] Yin, B. C.; Sun, Y. F.; Wang, C. Z.; Gai, Y. BJUT-3D large scale 3D face database and information processing. *Journal of Computer Research and Development* Vol. 46, No. 6, 1009–1018, 2009. (in Chinese)

[34] Gupta, S.; Castleman, K. R.; Markey, M. K.; Bovik, A. C. Texas 3D face recognition database. In: Proceedings of the IEEE Southwest Symposium on Image Analysis & Interpretation, 97–100, 2010.

[35] Vijayan, V.; Bowyer, K. W.; Flynn, P. J.; Huang, D.; Chen, L. M.; Hansen, M.; Ocegueda, O.; Shah, S. K.; Kakadiaris, I. A. Twins 3D face recognition challenge. In: Proceedings of the International Joint Conference on Biometrics, 1–7, 2011.

[36] Veltkamp, R.; van Jole, S.; Drira, H.; Amor, B.; Daoudi, M.; Li, H. B.; Chen, L. M.; Claes, P.; Smeets, D.; Hermans, J.; et al. SHREC'11 track: 3D face models retrieval. In: Proceedings of the 4th Eurographics Conference on 3D Object Retrieval, 89–95, 2011.

[37] Zhang, Y.; Guo, Z.; Lin, Z.; Zhang, H.; Zhang, C. The NPU multi-case Chinese 3D face database and information processing. *Chinese Journal of Electronics* Vol. 21, No. 2, 283–286, 2012.

[38] Zhang, X.; Yin, L. J.; Cohn, J. F.; Canavan, S.; Reale, M.; Horowitz, A.; Peng, L. A high-resolution spontaneous 3D dynamic facial expression database. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 1–6, 2013.

[39] Min, R.; Kose, N.; Dugelay, J. L. KinectFaceDB: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* Vol. 44, No. 11, 1534–1548, 2014.

[40] Zhang, J. J.; Huang, D.; Wang, Y. H.; Sun, J. Lock3DFace: A large-scale database of low-cost Kinect 3D faces. In: Proceedings of the International Conference on Biometrics, 1–8, 2016.

[41] Urbanová, P.; Ferková, Z.; Jandová, M.; Jurda, M.; Černý, D.; Sochor, J. Introducing the FIDENTIS 3D face database. *Anthropological Review* Vol. 81, No. 2, 202–223, 2018.

[42] Zulqarnain Gilani, S.; Mian, A. Learning from millions of 3D scans for large-scale 3D face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1896–1905, 2018.

[43] Cheng, S. Y.; Kotsia, I.; Pantic, M.; Zafeiriou, S. 4DFAB: A large scale 4D database for facial expression analysis and biometric applications. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5117–5126, 2018.

[44] Jia, S.; Li, X.; Hu, C. B.; Guo, G. D.; Xu, Z. Q. 3D face anti-spoofing with factorized bilinear coding. *arXiv preprint* arXiv:2005.06514, 2020.

[45] Ye, Y. P.; Song, Z.; Guo, J. G.; Qiao, Y. SIAT-3DFE: A high-resolution 3D facial expression dataset. *IEEE Access* Vol. 8, 48205–48211, 2020.

[46] Yang, H. T.; Zhu, H.; Wang, Y. R.; Huang, M. K.; Shen, Q.; Yang, R. G.; Cao, X. FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 598–607, 2020.

[47] Li, Q.; Dong, X. X.; Wang, W. N.; Shan, C. F. CAS-AIR-3D face: A low-quality, multi-modal and multi-pose 3D face database. In: Proceedings of the IEEE International Joint Conference on Biometrics, 1–8, 2021.

[48] Gilani, S. Z.; Mian, A. Towards large-scale 3D face recognition. In: Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, 1–8, 2016.

[49] Farkas, L. G. *Anthropometry of the Head and Face*. Raven Press, 1994.

[50] Blanz, V.; Vetter, T. A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 187–194, 1999.

[51] Kim, D.; Hernandez, M.; Choi, J.; Medioni, G. Deep 3D face identification. In: Proceedings of the IEEE International Joint Conference on Biometrics, 133–142, 2017.

[52] Zhang, Z. Y.; Da, F. P.; Yu, Y. Data-free point cloud network for 3D face recognition. *arXiv preprint* arXiv:1911.04731, 2019.

[53] Deng, J. K.; Cheng, S. Y.; Xue, N. N.; Zhou, Y. X.; Zafeiriou, S. UV-GAN: Adversarial facial UV map completion for pose-invariant face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7093–7102, 2018.

[54] Zhao, J.; Xiong, L.; Cheng, Y.; Cheng, Y.; Li, J.; Zhou, L.; Xu, Y.; Karlekar, J.; Pranata, S.; Shen, S.; et al. 3D-aided deep pose-invariant face recognition. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, 1184–1190, 2018.

[55] Shen, Y. J.; Luo, P.; Yan, J. J.; Wang, X. G.; Tang, X. O. FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 821–830, 2018.

[56] Zhang, X. Y.; Zhao, Y.; Zhang, H. Dual-discriminator GAN: A GAN way of profile face recognition. In: Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications, 162–166, 2020.

[57] Marriott, R. T.; Romdhani, S.; Chen, L. M. A 3D GAN for improved large-pose facial recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13440–13450, 2021.

[58] Luo, M. D.; Cao, J.; Ma, X.; Zhang, X. Y.; He, R. FA-GAN: Face augmentation GAN for deformation-invariant face recognition. *IEEE Transactions on Information Forensics and Security* Vol. 16, 2341–2355, 2021.

[59] Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv preprint* arXiv:1312.6203, 2013.

[60] Zhao, W.; Chellappa, R.; Phillips, P. J.; Rosenfeld, A. Face recognition. *ACM Computing Surveys* Vol. 35, No. 4, 399–458, 2003.

[61] Berretti, S.; del Bimbo, A.; Pala, P. 3D partial face matching using local shape descriptors. In: Proceedings of the Joint ACM Workshop on Human Gesture and Behavior Understanding, 65–71, 2011.

[62] Li, H. B.; Huang, D.; Lemaire, P.; Morvan, J. M.; Chen, L. M. Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities. In: Proceedings of the 18th IEEE International Conference on Image Processing, 3053–3056, 2011.

[63] Creusot, C.; Pears, N.; Austin, J. Automatic keypoint detection on 3D faces using a dictionary of local shapes. In: Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 204–211, 2011.

[64] Zhang, G. P.; Wang, Y. H. Robust 3D face recognition based on resolution invariant features. *Pattern Recognition Letters* Vol. 32, No. 7, 1009–1019, 2011.

[65] Inan, T.; Halici, U. 3-D face recognition with local shape descriptors. *IEEE Transactions on Information Forensics and Security* Vol. 7, No. 2, 577–587, 2012.

[66] Berretti, S.; del Bimbo, A.; Pala, P. Sparse matching of salient facial curves for recognition of 3-D faces with missing parts. *IEEE Transactions on Information Forensics and Security* Vol. 8, No. 2, 374–389, 2013.

[67] Li, X. L.; Da, F. P. Efficient 3D face recognition handling facial expression and hair occlusion. *Image and Vision Computing* Vol. 30, No. 9, 668–679, 2012.

[68] Ballihi, L.; Ben Amor, B.; Daoudi, M.; Srivastava, A.; Aboutajdine, D. Boosting 3-D-geometric features for efficient face recognition and gender classification. *IEEE Transactions on Information Forensics and Security* Vol. 7, No. 6, 1766–1779, 2012.

[69] Berretti, S.; Werghi, N.; del Bimbo, A.; Pala, P. Matching 3D face scans using interest points and local histogram descriptors. *Computers & Graphics* Vol. 37, No. 5, 509–525, 2013.

[70] Smeets, D.; Keustermans, J.; Vandermeulen, D.; Suetens, P. meshSIFT: Local surface features for 3D face recognition under expression variations and partial data. *Computer Vision and Image Understanding* Vol. 117, No. 2, 158–169, 2013.

[71] Creusot, C.; Pears, N.; Austin, J. A machine-learning approach to keypoint detection and landmarking on 3D meshes. *International Journal of Computer Vision* Vol. 102, Nos. 1–3, 146–179, 2013.

[72] Tang, H. L.; Yin, B. C.; Sun, Y. F.; Hu, Y. L. 3D face recognition using local binary patterns. *Signal Processing* Vol. 93, No. 8, 2190–2198, 2013.

[73] Lei, Y. J.; Bennamoun, M.; El-Sallam, A. A. An efficient 3D face recognition approach based on the fusion of novel local low-level features. *Pattern Recognition* Vol. 46, No. 1, 24–37, 2013.

[74] Elaiwat, S.; Bennamoun, M.; Boussaid, F.; El-Sallam, A. 3-D face recognition using curvelet local features. *IEEE Signal Processing Letters* Vol. 21, No. 2, 172–175, 2014.

[75] Drira, H.; Ben Amor, B.; Srivastava, A.; Daoudi, M.; Slama, R. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 35, No. 9, 2270–2283, 2013.

[76] Li, H. B.; Huang, D.; Morvan, J. M.; Chen, L. M.; Wang, Y. H. Expression-robust 3D face recognition *via* weighted sparse representation of multi-scale and multi-component local normal patterns. *Neurocomputing* Vol. 133, 179–193, 2014.

[77] Berretti, S.; Werghi, N.; Bimbo, A.; Pala, P. Selecting stable keypoints and local descriptors for person identification using 3D face scans. *The Visual Computer* Vol. 30, No. 11, 1275–1292, 2014.

[78] Lei, Y. J.; Bennamoun, M.; Hayat, M.; Guo, Y. L. An efficient 3D face recognition approach using local geometrical signatures. *Pattern Recognition* Vol. 47, No. 2, 509–524, 2014.

[79] Tabia, H.; Laga, H.; Picard, D.; Gosselin, P. H. Covariance descriptors for 3D shape matching and retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4185–4192, 2014.

[80] Vezzetti, E.; Marcolin, F.; Fracastoro, G. 3D face recognition: An automatic strategy based on geometrical descriptors and landmarks. *Robotics and Autonomous Systems* Vol. 62, No. 12, 1768–1776, 2014.

[81] Li, H. B.; Huang, D.; Morvan, J. M.; Wang, Y. H.; Chen, L. M. Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors. *International Journal of Computer Vision* Vol. 113, No. 2, 128–142, 2015.

[82] Elaiwat, S.; Bennamoun, M.; Boussaid, F.; El-Sallam, A. A curvelet-based approach for textured 3D face recognition. *Pattern Recognition* Vol. 48, No. 4, 1235–1246, 2015.

[83] Al-Osaimi, F. R. A novel multi-purpose matching representation of local 3D surfaces: A rotationally invariant, efficient, and highly discriminative approach with an adjustable sensitivity. *IEEE Transactions on Image Processing* Vol. 25, No. 2, 658–672, 2016.

[84] Ming, Y. Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. *Image and Vision Computing* Vol. 35, 14–22, 2015.

[85] Guo, Y. L.; Lei, Y. J.; Liu, L.; Wang, Y.; Bennamoun, M.; Sohel, F. EI3D: Expression-invariant 3D face recognition based on feature and shape matching. *Pattern Recognition Letters* Vol. 83, 403–412, 2016.

[86] Soltanpour, S.; Wu, Q. J. Multimodal 2D–3D face recognition using local descriptors: Pyramidal shape map and structural context. *IET Biometrics* Vol. 6, No. 1, 27–35, 2017.

[87] Lei, Y. J.; Guo, Y. L.; Hayat, M.; Bennamoun, M.; Zhou, X. Z. A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample. *Pattern Recognition* Vol. 52, 218–237, 2016.

[88] Emambakhsh, M.; Evans, A. Nasal patches and curves for expression-robust 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 39, No. 5, 995–1007, 2017.

[89] Werghi, N.; Tortorici, C.; Berretti, S.; Del Bimbo, A. Boosting 3D LBP-based face recognition by fusing shape and texture descriptors on the mesh. *IEEE Transactions on Information Forensics and Security* Vol. 11, No. 5, 964–979, 2016.

[90] Hariri, W.; Tabia, H.; Farah, N.; Benouareth, A.; Declercq, D. 3D face recognition using covariance based descriptors. *Pattern Recognition Letters* Vol. 78, 1–7, 2016.

[91] Soltanpour, S.; Jonathan Wu, Q. M. High-order local normal derivative pattern (LNDP) for 3D face recognition. In: Proceedings of the IEEE International Conference on Image Processing, 2811–2815, 2017.

[92] Deng, X.; Da, F. P.; Shao, H. J. Efficient 3D face recognition using local covariance descriptor and Riemannian kernel sparse coding. *Computers & Electrical Engineering* Vol. 62, 81–91, 2017.

[93] Abbad, A.; Abbad, K.; Tairi, H. 3D face recognition: Multi-scale strategy based on geometric and local descriptors. *Computers & Electrical Engineering* Vol. 70, 525–537, 2018.

[94] Soltanpour, S.; Wu, Q. M. J. Weighted extreme sparse classifier and local derivative pattern for 3D face recognition. *IEEE Transactions on Image Processing* Vol. 28, No. 6, 3020–3033, 2019.

[95] Shi, L. L.; Wang, X.; Shen, Y. L. Research on 3D face recognition method based on LBP and SVM. *Optik* Vol. 220, 165157, 2020.

[96] Samir, C.; Srivastava, A.; Daoudi, M.; Klassen, E. An intrinsic framework for analysis of facial surfaces. *International Journal of Computer Vision* Vol. 82, No. 1, 80–95, 2009.

[97] Samir, C.; Srivastava, A.; Daoudi, M. Three-dimensional face recognition using shapes of facial curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 28, No. 11, 1858–1863, 2006.

[98] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* Vol. 60, No. 2, 91–110, 2004.

[99] Deng, X.; Da, F.; Shao, H. J.; Jiang, Y. T. A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors. *Computers & Electrical Engineering* Vol. 85, 106700, 2020.

[100] Vezzetti, E.; Marcolin, F.; Tornincasa, S.; Ulrich, L.; Dagnes, N. 3D geometry-based automatic landmark localization in presence of facial occlusions. *Multimedia Tools and Applications* Vol. 77, No. 11, 14177–14205, 2018.

[101] Drira, H.; Benamor, B.; Daoudi, M.; Srivastava, A. Pose and expression-invariant 3D face recognition using elastic radial curves. In: Proceedings of the British Machine Vision Conference, 1–11, 2010.

[102] Freund, Y.; Schapire, R. E. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence* Vol. 14, No. 5, 771–780, 1999.

[103] Aubry, M.; Schlickewei, U.; Cremers, D. The wave kernel signature: A quantum mechanical approach to shape analysis. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, 1626–1633, 2011.

[104] Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 24, No. 7, 971–987, 2002.

[105] Yu, Y.; Da, F. P.; Guo, Y. F. Sparse ICP with resampling and denoising for 3D face verification. *IEEE Transactions on Information Forensics and Security* Vol. 14, No. 7, 1917–1927, 2019.

[106] Spreeuwers, L. Fast and accurate 3D face recognition. *International Journal of Computer Vision* Vol. 93, No. 3, 389–414, 2011.

[107] Ocegueda, O.; Passalis, G.; Theoharis, T.; Shah, S. K.; Kakadiaris, I. A. UR3D-C: Linear dimensionality reduction for efficient 3D face recognition. In: Proceedings of the International Joint Conference on Biometrics, 1–6, 2011.

[108] Ming, Y.; Ruan, Q. Q. Robust sparse bounding sphere for 3D face recognition. *Image and Vision Computing* Vol. 30, No. 8, 524–534, 2012.

[109] Liu, P. J.; Wang, Y. H.; Huang, D.; Zhang, Z. X.; Chen, L. M. Learning the spherical harmonic features for 3-D face recognition. *IEEE Transactions on Image Processing* Vol. 22, No. 3, 914–925, 2013.

[110] Taghizadegan, Y.; Ghassemian, H.; Naser-Moghaddasi, M. 3D face recognition method using 2DPCA-Euclidean distance classification. *ACEEE International Journal on Control System and Instrumentation* Vol. 3, No. 1, 1–5, 2012.

[111] Mohammadzade, H.; Hatzinakos, D. Iterative closest normal point for 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 35, No. 2, 381–397, 2013.

[112] Ming, Y. Rigid-area orthogonal spectral regression for efficient 3D face recognition. *Neurocomputing* Vol. 129, 445–457, 2014.

[113] Ratyal, N. I.; Ahmad Taj, I.; Bajwa, U. I.; Sajid, M. 3D face recognition based on pose and expression invariant alignment. *Computers & Electrical Engineering* Vol. 46, 241–255, 2015.

[114] Tang, Y. H.; Sun, X.; Huang, D.; Morvan, J. M.; Wang, Y. H.; Chen, L. M. 3D face recognition with asymptotic cones based principal curvatures. In: Proceedings of the International Conference on Biometrics, 466–472, 2015.

[115] Gilani, S. Z.; Mian, A.; Eastwood, P. Deep, dense and accurate 3D face correspondence for generating population specific deformable models. *Pattern Recognition* Vol. 69, 238–250, 2017.

[116] Peter, M.; Minoi, J. L.; Hipiny, I. H. M. 3D face recognition using kernel-based PCA approach. In: *Computational Science and Technology. Lecture Notes in Electrical Engineering, Vol. 481*. Alfred, R.; Lim, Y.; Ibrahim, A.; Anthony, P. Eds. Springer Singapore, 77–86, 2019.

[117] Passalis, G.; Perakis, P.; Theoharis, T.; Kakadiaris, I. A. Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 33, No. 10, 1938–1951, 2011.

[118] Huang, D.; Ardabilian, M.; Wang, Y. H.; Chen, L. M. 3-D face recognition using eLBP-based facial description and local feature hybrid matching. *IEEE Transactions on Information Forensics and Security* Vol. 7, No. 5, 1551–1565, 2012.

[119] Alyüz, N.; Gökberk, B.; Spreeuwers, L.; Veldhuis, R.; Akarun, L. Robust 3D face recognition in the presence of realistic occlusions. In: Proceedings of the 5th IAPR International Conference on Biometrics, 111–118, 2012.

[120] Fadaifard, H.; Wolberg, G.; Haralick, R. Multiscale 3D feature extraction and matching with an application to 3D face recognition. *Graphical Models* Vol. 75, No. 4, 157–176, 2013.

[121] Alyuz, N.; Gokberk, B.; Akarun, L. 3-D face recognition under occlusion using masked projection. *IEEE Transactions on Information Forensics and Security* Vol. 8, No. 5, 789–802, 2013.

[122] Bagchi, P.; Bhattacharjee, D.; Nasipuri, M. Robust 3D face recognition in presence of pose and partial occlusions or missing parts. *arXiv preprint* arXiv:1408.3709, 2014.

[123] Bagchi, P.; Bhattacharjee, D.; Nasipuri, M. 3D Face Recognition using surface normals. In: Proceedings of the TENCON 2015 - 2015 IEEE Region 10 Conference, 1–6, 2015.

[124] Liang, Y.; Zhang, Y.; Zeng, X. X. Pose-invariant 3D face recognition using half face. *Signal Processing: Image Communication* Vol. 57, 84–90, 2017.

[125] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* Vol. 521, No. 7553, 436–444, 2015.

[126] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1701–1708, 2014.

[127] Sun, Y.; Chen, Y. H.; Wang, X. G.; Tang, X. O. Deep learning face representation by joint identification-verification. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2, 1988–1996, 2014.

[128] Sun, Y.; Wang, X. G.; Tang, X. O. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1891–1898, 2014.

[129] Sun, Y.; Wang, X. G.; Tang, X. O. Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2892–2900, 2015.

[130] Sun, Y.; Liang, D.; Wang, X. G.; Tang, X. O. DeepID3: Face recognition with very deep neural networks. *arXiv preprint* arXiv:1502.00873, 2015.

[131] Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815–823, 2015.

[132] Charles, R. Q.; Hao, S.; Mo, K. C.; Guibas, L. J. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 77–85, 2017.

[133] Qi, C. R.; Yi, L.; Su, H.; Guibas, L. J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing System, 5105–5114, 2017.

[134] Ding, Y. Q.; Li, N. Y.; Young, S. S.; Ye, J. W. Efficient 3D face recognition in uncontrolled environment. In: *Advances in Visual Computing. Lecture Notes in Computer Science, Vol. 11844*. Springer Cham, 430–443, 2019.

[135] Lin, S. S.; Liu, F.; Liu, Y. H.; Shen, L. L. Local feature tensor based deep learning for 3D face recognition. In: Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition, 1–5, 2019.

[136] Tan, Y.; Lin, H. X.; Xiao, Z. L.; Ding, S. Y.; Chao, H. Y. Face recognition from sequential sparse 3D data via deep registration. In: Proceedings of the International Conference on Biometrics, 1–8, 2019.

[137] Olivetti, E. C.; Ferretti, J.; Cirrincione, G.; Nonis, F.; Tornincasa, S.; Marcolin, F. Deep CNN for 3D face recognition. In: *Design Tools and Methods in Industrial Engineering. Lecture Notes in Mechanical Engineering*. Rizzi, C.; Andrisano, A. O.; Leali, F.; Gherardini, F.; Pini, F.; Vergnano, A. Eds. Springer Cham, 665–674, 2020.

[138] Xu, K. M.; Wang, X. M.; Hu, Z. H.; Zhang, Z. H. 3D face recognition based on twin neural network combining deep map and texture. In: Proceedings of the IEEE 19th International Conference on Communication Technology, 1665–1668, 2019.

[139] Feng, J. Y.; Guo, Q.; Guan, Y. D.; Wu, M. D.; Zhang, X. R.; Ti, C. L. 3D face recognition method based on deep convolutional neural network In: *Smart Innovations in Communication and Computational Sciences. Advances in Intelligent Systems and Computing, Vol. 670*. Panigrahi, B.; Trivedi, M.; Mishra, K.; Tiwari, S.; Singh, P. Eds. Springer Singapore, 123–130, 2019.

[140] Mu, G. D.; Huang, D.; Hu, G. S.; Sun, J.; Wang, Y. H. Led3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5766–5775, 2019.

[141] Dutta, K.; Bhattacharjee, D.; Nasipuri, M. SpPCANet: A simple deep learning-based feature extraction approach for 3D face recognition. *Multimedia Tools and Applications* Vol. 79, Nos. 41–42, 31329–31352, 2020.

[142] Cao, C. Q.; Swash, M. R.; Meng, H. Y. Reliable holoscopic 3D face recognition. In: Proceedings of the 7th International Conference on Signal Processing and Integrated Networks, 696–701, 2020.

[143] Lin, S. S.; Jiang, C. Y.; Liu, F.; Shen, L. L. High quality facial data synthesis and fusion for 3D low-quality face recognition. In: Proceedings of the IEEE International Joint Conference on Biometrics, 1–8, 2021.

[144] Chiu, M. T.; Cheng, H. Y.; Wang, C. Y.; Lai, S. H. High-accuracy RGB-D face recognition via segmentation-aware face depth estimation and mask-guided attention network. In: Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition, 1–8, 2021.

[145] Zhang, Z. Y.; Da, F. P.; Yu, Y. Learning directly from synthetic point clouds for "in-the-wild" 3D face recognition. *Pattern Recognition* Vol. 123, 108394, 2022.

[146] Bhople, A. R.; Shrivastava, A. M.; Prakash, S. Point cloud based deep convolutional neural network for 3D face recognition. *Multimedia Tools and Applications* Vol. 80, No. 20, 30237–30259, 2021.

[147] Bhople, A. R.; Prakash, S. Learning similarity and dissimilarity in 3D faces with triplet network. *Multimedia Tools and Applications* Vol. 80, Nos. 28–29, 35973–35991, 2021.

[148] Jiang, C. Y.; Lin, S. S.; Chen, W.; Liu, F.; Shen, L. L. PointFace: Point set based feature learning for 3D face recognition. In: Proceedings of the IEEE International Joint Conference on Biometrics, 1–8, 2021.

[149] Papadopoulos, K.; Kacem, A.; Shabayek, A.; Aouada, D. Face-GCN: A graph convolutional network for 3D dynamic face identification/recognition. *arXiv preprint* arXiv:2104.09145, 2021.

[150] Li, B. Y. L.; Mian, A. S.; Liu, W. Q.; Krishna, A. Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In: Proceedings of the IEEE Workshop on Applications of Computer Vision, 186–192, 2013.

[151] Wang, F.; Cheng, J.; Liu, W. Y.; Liu, H. J. Additive margin softmax for face verification. *IEEE Signal Processing Letters* Vol. 25, No. 7, 926–930, 2018.

[152] Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* Vol. 20, No. 3, 273–297, 1995.

[153] Isola, P.; Zhu, J. Y.; Zhou, T. H.; Efros, A. A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5967–5976, 2017.

[154] Jiang, L.; Zhang, J. Y.; Deng, B. L. Robust RGB-D face recognition using attribute-aware loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 42, No. 10, 2552–2566, 2020.

[155] Paysan, P.; Knothe, R.; Amberg, B.; Romdhani, S.; Vetter, T. A 3D face model for pose and illumination invariant face recognition. In: Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, 296–301, 2009.

[156] Cao, C.; Weng, Y. L.; Zhou, S.; Tong, Y. Y.; Zhou, K. FaceWarehouse: A 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics* Vol. 20, No. 3, 413–425, 2014.

[157] Castellani, U.; Bartoli, A. 3D shape registration. In: *3D Imaging, Analysis and Applications*. Liu, Y.; Pears, N.; Rosin, P. L.; Huber, P. Eds. Springer Cham, 353–411, 2020.

[158] D'Errico, J. Surface fitting using gridfit. MATLAB Central File Exchange. 2005. Available at https://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit.

[159] Ciravegna, G.; Cirrincione, G.; Marcolin, F.; Barbiero, P.; Dagnes, N.; Piccolo, E. Assessing discriminating capability of geometrical descriptors for 3D face recognition by using the GH-EXIN neural network. In: *Neural Approaches to Dynamics of Signal Exchanges. Smart Innovation, Systems and Technologies, Vol. 151*. Esposito, A.; Faundez-Zanuy, M.; Morabito, F.; Pasero, E. Eds. Springer Singapore, 223–233, 2020.

[160] Lüthi, M.; Gerig, T.; Jud, C.; Vetter, T. Gaussian process morphable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 40, No. 8, 1860–1873, 2017.

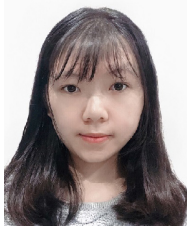
[161] Gilani, S. Z.; Mian, A.; Shafait, F.; Reid, I. Dense 3D face correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 40, No. 7, 1584–1598, 2018.

[162] El Rahman Shabayek, A.; Aouada, D.; Cherenkova, K.; Gusev, G.; Ottersten, B. 3D deformation signature for dynamic face recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2138–2142, 2020.

[163] Smith, M.; Smith, L.; Huang, N.; Hansen, M.; Smith, M. Deep 3D face recognition using 3D data augmentation and transfer learning. In: Proceedings of the 16th International Conference on Machine Learning and Data Mining, 209–218, 2020.

[164] Dou, P. F.; Shah, S. K.; Kakadiaris, I. A. End-to-end 3D face reconstruction with deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1503–1512, 2017.

[165] Richardson, E.; Sela, M. T.; Kimmel, R. 3D face reconstruction by learning from synthetic data. In: Proceedings of the 4th International Conference on 3D Vision, 460–469, 2016.



Yaping Jing received her bachelor degree (Hons) in information technology from Deakin University, Australia in 2016. She is currently a Ph.D. candidate in the School of Information Technology at Deakin University. Her research interests include 3D face recognition, 3D data processing, and machine learning.

Xuequan Lu is a lecturer (assistant professor) in Deakin University, Australia. He spent more than two years working as a research fellow in Singapore. Prior to that, he received his Ph.D. degree from Zhejiang University, China in 2016. His research interests mainly fall into the category of visual data computing, for example, geometry modeling, processing and analysis, animation, simulation, 2D data processing and analysis.

Shang Gao received her Ph.D. degree in computer science from Northeastern University, China in 2000. She is currently a senior lecturer in the School of Information Technology, Deakin University. Her current research interests include cybersecurity, cloud computing, and machine learning.