3D Face Recognition: A Survey

Yaping Jing, Xuequan Lu, *Member, IEEE,* and Shang Gao

***Abstract*— Face recognition is one of the most studied research topics in the community. In recent years, the research on face recognition has shifted to using 3D facial surfaces, as more discriminating features can be represented by the 3D geometric information. This survey focuses on reviewing the 3D face recognition techniques developed in the past ten years which are generally categorized into conventional methods and deep learning methods. The categorized techniques are evaluated using detailed descriptions of the representative works. The advantages and disadvantages of the techniques are summarized in terms of accuracy, complexity and robustness to face variation (expression, pose and occlusions, etc). The main contribution of this survey is that it comprehensively covers both conventional methods and deep learning methods on 3D face recognition. In addition, a review of available 3D face databases is provided, along with the discussion of future research challenges and directions.**

***Index Terms*— 3D face recognition, 3D face database, Survey, Deep Learning, Local feature, Global feature**

I. Introduction

Face recognition has become a commonly used biometric technology, which is widely applied in public records, authentication, security, intelligence and many other vigilance systems [1]. During the past decades, many 2D face recognition techniques have achieved high performance under controlled environments. The accuracy of 2D face recognition has been greatly enhanced especially after the emergence of deep learning. However, these techniques are still challenged by the intrinsic limitations of 2D images, such as illumination, pose, expression, occlusion, disguise, time delay and image quality [2]. 3D face recognition may outperform 2D face recognition [3] with greater recognition accuracy and robustness, as it is less sensitive to pose, illumination, and expression [4]. Furthermore, richer geometric information on 3D face can provide more discriminative features for face recognition. Thus, 3D face recognition has become an active research topic in recent years.

In 3D face recognition, 3D face models are normally used for training and testing purposes. Compared with 2D images, 3D face models contain more shape information. These rigid features can help face recognition systems overcome the inherent defects and drawbacks of 2D face recognition, for example, the facial expression, occlusion, and pose variations. Furthermore, a 3D model is relatively unchanged in terms of scaling, rotation, and illumination [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 performance.

One of the main challenges of 3D face recognition is the acquisition of 3D images as it cannot be accomplished by crawling the Web like how 2D face images are collected. It requires special hardware equipment instead. According to the technologies used, it can be broadly divided into active acquisition and passive acquisition [6]. An active collection system actively emits invisible light (e.g. infrared laser beam) to illuminate the target face and obtain the shape features of the target by measuring the reflectivity. A passive acquisition system consists of several cameras placed separately from each other. It matches points observed from other cameras and calculates the exact 3D position of the matched point. The 3D surface is formed by a set of matched points. Since 2000, many researchers have begun to conduct an assessment of 3D face recognition algorithms on large-scale 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. In this paper, we also summarize the existing public 3D face databases and particularly their data augmentation methods when reviewing these recognition technologies.

There were relevant surveys conducted by researchers from different perspectives. In 2006, Bowyer et al. [3] reviewed the research trends in 3D face recognition. Abate et al. [11] summarized the associated literature up to year 2007. Smeets et al. (2012) [12] studied various algorithms for expression invariant 3D face recognition and evaluated the complexity of existing 3D face databases. Followed by that, Zhou et al. [2] categorized face recognition algorithms into single-modal and multi-modal ones in 2014. Patil et al. (2015) [1] studied the 3D face recognition techniques that comprehensively covered the conventional methods. Recently, [13] and [6] both presented a review of the 3D face recognition algorithms, but only a few deep learning-based methods were covered. [14] and [15] reviewed the deep learning-based face recognition methods in 2018, but the focus was mainly on 2D face recognition. In this paper, we focus on 3D face recognition. Compared with the existing literature, the main contributions of our work are summarized as follows:

- To the best of our knowledge, this is the first survey paper that comprehensively covers conventional methods and deep learning-based methods for 3D face recognition.
- Different from the existing surveys, it pays special

Y. Jing, X. Lu and S. Gao are all with the School of Information Technology, Deakin University, Waurn Ponds, VIC, Australia. Corresponding author: X. Lu, S. Gao. E-mails: {jingyap, xuequan.lu, shang.gao}@deakin.edu.au.

attention to deep learning-based 3D face recognition methods.
- It covers the latest and most advanced development in 3D face recognition, providing a clear progress chart for 3D face recognition.
- It provides a comprehensive comparison of existing methods on the available datasets, and it suggests future research challenges and directions.

According to the feature extraction methods adopted, 3D face recognition techniques can be divided into two categories: conventional method and deep learning-based method (Fig. 1). The conventional methods always use traditional algorithms to extract face features, e.g. Iterative closest point (ICP), principle component analysis (PCA), linear and nonlinear algorithms. They can be further divided into three types: local feature-based, holistic-based and hybrid. As for the deep learning-based methods, nearly all deep learning-based methods use pre-trained networks and then fine-tune these networks with the converted data (e.g. 2D images 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]. The structure of this paper is as follows. Section 2 introduces the widely used 3D face databases/datasets. Section 3 and 4 review conventional 3D face recognition methods and deep learning-based methods, respectively. Section 5 compares these methods and discuss future research directions, followed by a conclusion in Section 6.

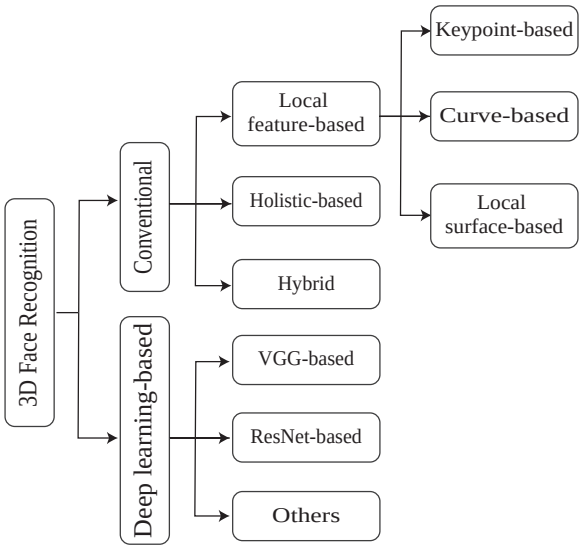
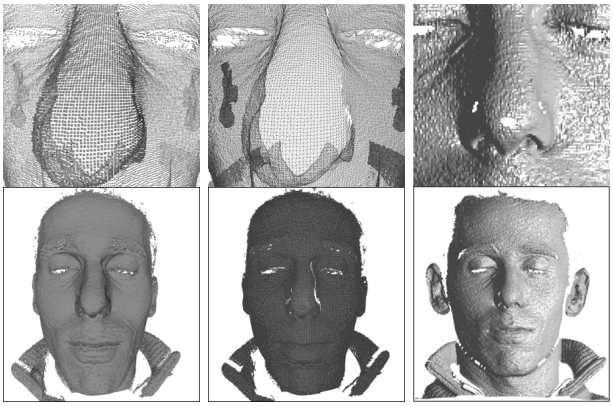


Fig. 1: A taxonomy of 3D face recognition methods.

II. 3D FACE DATABASE

Large-scale 3D face databases/datasets are essential for the development of 3D face recognition. They are used to train the feature extraction algorithms and evaluate their performance. To meet this demand, many research institutions and researchers have established various 3D face databases. Table I enlists the currently prominent 3D face databases and compares the data formats, the number of identities, image variations (e.g. expression, pose, and occlusion), and the scanner devices. There are four different 3D data formats: point cloud (Fig. 2(a)), meshes (Fig. 2(b)), range image (Fig. 2(c)) or depth maps, and 3D video; and two types of acquisition scanner devices: laser-based and stereo-based. Laser-based active collection systems use structured light scanners (e.g. Microsoft Kinect) or laser scanners (e.g. Minolta vivid scanner). Stereo-based devices are used by a passive acquisition system, such as Bumblebee XB3.



(a) Point cloud (b) 3D mesh (c) Range image

Fig. 2: 3D face data representations [20].

Before 2004, there were few public 3D face databases. Some representatives include **3DRMA** [21], **FSU** [22] and **GavabDB** [23].The **GavabDB** database consists of 61 individuals, aged between 18 and 40. Each identity has 3 frontal images with different expressions and 4 rotating images without expressions [23]. In 2005, **FRGC V2.0** database was designed to improve the performance of face recognition algorithms, which had a huge impact on the development of 3D face recognition [9]. So far, it is still used as a standard reference database (SRD) for evaluating the performance of 3D face recognition algorithms. In the same year, another important database **UND** (the University of Notre Dame) face database was released, where each identity has only one 3D image and multiple 2D images [24].

From 2006 to 2010, there were more databases created. The largest one is **ND-2006** which is a superset of FRGC V2. It contains 13,450 images and 888 persons with as many as 63 images per identity [29]. The second largest is **UoY** database, which consists of more than 5,000 models (350 people) owned by the University of York (UK) [31]. The **CASIA** and **Bosphorus** database are similar in size, close to 5,000 images. The **CASIA** database was collected by using the non-contact 3D digitizer Minolta Vivid 910 in 2004 and contains 4,059 images of 123 objects [27]. It not only considers the individual variation of expressions, poses, and illumination but also introduces the combined changes of different expressions in different poses. **Bosphorus** has 381 individuals and the most expression and posture

TABLE I: 3D face databases.

Name/Reference	Year	Data type	IDs	Scans	Texture	Expression	Pose	Occlusion	Scanner
3DRMA[21]	2000	Mesh	120	720	Yes	-	Slight left/right, up/down	-	Structured light
FSU[22]	2003	Mesh	37	222	No	-	-	-	Laser
GavabDB[23]	2004	Mesh	61	427	No	Neutral, smile, accentuated laugh	$\pm 30^{\circ}$	-	Laser
FRGC v2[9]	2005	Range image	466	4,007	Yes	neutral, smiling	$\pm 15^{\circ}$	-	Laser
UND[24]	2005	Range image	275	670	Yes	-	$\pm 45^{\circ}$, $\pm 60^{\circ}$	-	Laser
ZJU-3DFED[25]	2006	Mesh	40	360	No	Neutral, smile, surprise, sad	-	-	Structured light
BU3D-FE[26]	2006	Mesh	100	2,500	Yes	Anger, happiness, sadness, surprise, disgust, fear	-	-	Stereo
CASIA[27]	2006	Range image	123	4,059	No	Neutral, smile, eyes-closed, anger, laugh, surprise	$\pm 90^{\circ}$	-	Laser
FRAV3D[28]	2006	Mesh	105	1,696	Yes	Neutral, smile, open mouth, gesture	Up and down, Y-axis turn, z-axis turn	-	Laser
ND-2006[29]	2007	Range image	888	13,450	Yes	Neutral, happiness, sadness, surprise, disgust, and other	$\pm 15^{\circ}$	-	Laser
Bosphorus[30]	2008	Point cloud	105	4,666	Yes	34	3 yaw, pitch, cross rotations	4 types	Stereo
UoY[31]	2008	Mesh	350	5,000	Yes	Neutral, eyes closed, eyebrows raised, happy, anger	Frontal, up, down	-	Stereo
SHREC08[32]	2008	Range image	61	427	No	Smile, laugh and arbitrary expressions	Front, up, down	-	-
BJUT-3D[33]	2009	Mesh	500	1,200	Yes	-	-	-	Laser
Texas-3D[34]	2010	Range image	118	1,149	Yes	Smiling, talking faces with open/closed mouths & eyes	Frontal, $\pm 10^{\circ}$	-	Stereo
UMBDB[20]	2011	Range image	143	1,473	Yes	Neutral, smiling, angry, bored	Frontal	7 types	Laser
3D-TEC[35]	2011	Range image	214	428	Yes	Neutral, smiling	Frontal	-	Laser
SHREC11[36]	2011	Range image	130	780	No	-	5 directions	-	Laser
NPU3D[37]	2012	Mesh	300	10,500	No	9	14	4	Laser
BU4D-FE[38]	2013	3D video	101	60,600	Yes	-	-	-	Stereo
KinectFaceDB[39]	2014	Range image	52	936	Yes	Neutral, smiling, mouth open	Left, right	Multiple	Kinect
Lock3DFace[40]	2016	Range image	509	5,711	Yes	Happiness, anger, sadness, surprise, fear, disgust	$\pm 90^{\circ}$	Randomly cover-up	Kinect
F3D-FD[41]	2018	Range image	2,476	-	Yes	-	Semi-lateral with ear	Half face	Stereo
LS3DFace[42]	2018	Point cloud	1,853	31,860	Yes	-	-	-	-
WFFD[43]	2020	Videos	241	285	Yes	-	-	-	-
SIAT-3DFE[44]	2020	3D	500	8,000	Yes	16	-	2	Structured light
FaceScape[45]	2020	Videos	938	18,760	Yes	20	-	-	68 DSLR cameras

changes. It provides manual marking of 24 facial landmarks for each scanned image, such as nose tip, chin middle, eye corners [30]. Another database which includes manual landmarks is **Texas-3D**. In Texas-3D, these 3D images have been preprocessed and 25 manually landmarks are provided. Therefore, It provides a good option for researchers to focus specifically on developing 3D face recognition algorithms, without considering the initial preprocessing of 3D images [34]. The **BU3D-FE** (Binghamton University 3D Facial Expression) is a database specially developed for 3D facial expression recognition. It contains 100 identities with 6 expression types: anger, happiness, sadness, surprise, disgust and fear [26]. For the **FRAV3D** database, 81 males and 24 females were involved, and 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** is one of the largest Chinese 3D face databases which includes 1,200 Chinese 3D face images [33]. The two smallest databases are **ZJU-3DFED** and **SHREC08**. The **ZJU-3DFED** database consists of 40 identities and 9 scans with four different kinds of expressions for each identity [25]. The **SHREC08** database consists of 61 people with 7 scans

for each [32].

In the following five years after 2010, there were six remarkable databases created. The **UMBDB** database is an excellent database for testing the occlusion variance 3D face recognition algorithms, which contains 578 occlusion acquisitions [20]. **3D-TEC** (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 to promote the performance of 3D face recognition technology. The **SHREC11** is based on a new collection of 130 masks with 6 3D face scans [36]. In addition to **BJUT-3D**, Northwestern Polytechnical University 3D (**NPU3D**) is another large-scale Chinese 3D face database, composed of 10,500 3D face data, corresponding to 300 individuals[37]. The **BU4D-FE** is a 3D video database that records the spontaneous expressions of various young people by completing 8 emotional expression elicitation tasks [38]. The **KinectFaceDB** is the first publicly available face database based on the Kinect sensor and consists of four data modalities (2D, 2.5D, 3D, and video-based) [39].

Recently, another large-scale 3D face database **Lock3DFace** was released. It is based on Kinect and contains various variations in expressions, poses, time-lapse, and occlusions [40]. **F3D-FD** is a large dataset which has the most individuals (2,476). For each individual, it includes partial 3D scans from the frontal and two semi-later views, and a one-piece face with lateral parts (including ears, earless, with landmarks) [41]. The **LS3DFace** is the largest dataset so far, including 31,860 3D face scans of 1,853 identities. It is composed of multiple challenging public datasets, including FRGC v2, BU3D-FE, Bosphorus, GavabDB, Texas-3D, BU4D-FE, CASIA, UMBDB, 3D-TEC and ND-2006 [42]. The large-scale Wax Figure Face Database (**WFFD**) is designed to address the vulnerabilities in the existing 3D facial spoofing database and promote the research of 3D facial presentation attack detection [43]. This database includes photo-based and video-based data. We only detail the video information in Table I. **SIAT-3DFE** is a 3D facial expression dataset in which every identity has 16 facial expressions including natural, happiness, sadness, surprise, several exaggerated expressions (open mouth, frown, etc.), and two occluded 3D models [44]. Another recent database is **FaceScape**, which consists of 18,760 textured 3D models with pore-level facial geometry [45].

It is well known that the performance of 3D face recognition algorithms could change on different 3D face databases. Increasing the gallery size could degrade the performance of face recognition [46]. Although some algorithms have achieved good results on these existing 3D face databases, they still cannot be used in the real world due to more uncontrolled conditions in the real world. The establishment of large-scale 3D face databases to simulate real situations are very essential to facilitate the research of 3D face recognition. In addition, collecting 3D face data is a very time-consuming and source-demanding task. Research on large dataset generating algorithms would be one of the future works.

III. Conventional methods

In a conventional 3D face recognition system, there are two main phases: training and testing, as shown in Fig. 3. In the training phase, 3D face data is required to generate a feature gallery. Facial features are obtained through a data preprocessing and feature extraction model before being saved in the feature gallery. In the testing phase, a probe is acquired as the target face, and the same data preprocessing and feature extraction process as the training phase is performed. Face recognition is a matching process. The feature vectors of the target face are compared with the feature vectors stored in the feature gallery. The gallery is scanned and returns the face which has the closest matching distance. If the distance is lower than a predefined threshold, the target face is marked as recognized, otherwise, it fails. Thus, a face recognition process contains three core steps: data preprocessing, feature extraction and face matching. All of them can influence the performance of recognition.

A. Data Preprocessing and matching

In most situations, the acquired raw 3D face data cannot be directly used as the input for feature extraction systems as it may contain redundant information [6], for example, hair, neck, and background context. This information will influence the accuracy of recognition. Thus, the 3D data is usually preprocessed before passing into a feature extraction model. In general, the data preprocessing could include three main parts: facial landmarks detection and orientation, data segmentation and face registration. Facial landmarks are a set of keypoints defined by anthropometric studies [47], which can be used to automatically localize and register a face. Some databases already provide the landmarks of a face image. Data segmentation is the process of utilizing facial landmarks such as the nose tip and eye corners, to segment the facial surface [47]. This process is always used for local-based methods, which extract identifiable facial parts like the nose and eyes part for feature extraction. As an essential step before feature extraction and matching, face registration is to convert the target surface (entire face or face parts) to align with the training surface in the gallery.

After extracting the feature vectors from the original face, comes the most important part - face matching. The distances between the target face and the stored features in gallery are calculated. The common metrics include the Euclidean distance, Hausdorff distance, and angular Mahalanobis distance.

According to Zhao et al. [48] and our review on the last-decade literature, conventional face recognition algorithms can be classified into three types based on their feature extraction approaches: local feature-based, holistic-based, and hybrid, as shown in Fig. 1. Local-based approaches mainly focus on the local facial features such as nose and eyes [48].In contrast to the local-based methods, the holistic-based approaches use the entire face to generate feature vectors for feature classification. Hybrid methods use both local and global facial features.

For the local-based methods, fusion schemes are used to improve accuracy. There are five fusion schemes: sensor level

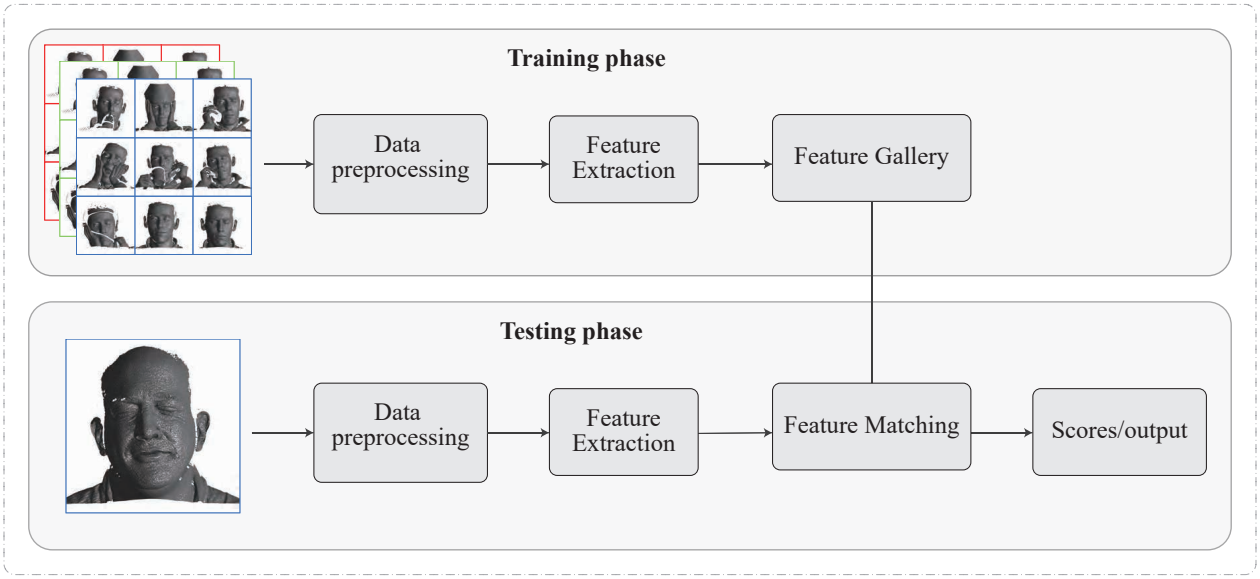


Fig. 3: The pipeline of 3D face recognition.

fusion, feature level fusion, rank level fusion, decision level fusion, and score level fusion [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 a single object; for rank level fusion, ranks are assigned to gallery images based on a descending sequence of confidence; score level fusion is a combination of the 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 methods are discussed below.

B. Local feature-based methods

In the last decade, many local feature-based approaches were built, where local feature descriptors were used to describe the 3D local facial information. Table II enlists the remarkable 3D local-based methods and summarizes their important details. According to [13], these methods can be classified into three different types based on the descriptors: keypoint-based, curve-based and local surface-based. For the keypoint-based methods, a set of 3D keypoints are detected based on the face geometric information and used to build feature descriptors by calculating relationships between these keypoints; The curve-based methods use a set of curves on one face surface as features vectors; The local surface-based methods extract features from some regions of face surface [13].

Keypoint-based: A keypoint-based method has two important steps: keypoint detection and feature descriptor construction [13]. It uses a set of keypoints and their geometric relationships to represent facial features. Therefore, it can partially process a face image with missing parts or occlusions. However, due to using a large number of keypoints, it also involves higher computational cost. The most effective keypoint selection is also crucial for creating effective feature vectors.

One of the most commonly used keypoint detectors is Scale Invariant Feature Transformation (SIFT) [49]. For example, [50] used SIFT to detect relevant keypoints of a 3D depth image, where local shape descriptors were adopted to measure the changes of face depth in the keypoints neighborhood. In [51], SIFT descriptors were applied to 2D matrices including shape index, curvedness, Gaussian and mean curvature values generated from 3D face data to obtain feature vectors. [52] used SIFT keypoint detection on pyramidal shape maps to obtain 3D geometric information and combine it with 2D keypoints. However, this SIFT method is sensitive to pose changes. [53] used a 3D point cloud registration algorithm combining with local features to achieve both pose and expression invariance. Later, a Keypoint-based Multiple Triangle Statistics (KMTS) method was proposed by [54] to address partial facial data, pose and large facial expression variations. Recently, SIFT was also used to detect keypoints in [55], which used local covariance descriptors and Riemann kernel sparse coding to improve the accuracy of 3D face recognition. The accuracy was further improved in [56].

In order to improve the robustness to large occlusions or poses, SIFT keypoint detection is directly used for 3D mesh data. The extension of SIFT for 3D mesh is called MeshSIFT [57]. In [57], salient points on 3D face surface are first detected as extreme values in a scale space, then an orientation is assigned to these points. A feature vector is used to describe them by concatenating the histograms of slant angles and shape indices. Before this approach was applied, [58] 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. [59] 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 Ballihi et al. ([60], [61]). They first used the meshDOG keypoint detector and local geometric histogram (GH) descriptor to extract features, then selected the most effective feature based on the analysis of the optimal scale, distribution and clustering of keypoints, and the features of the local descriptor. Recently, [62] exploited a curvelet-based multimodal keypoint detector and local surface descriptor that can extracts both texture and 3D local features. It reduces the computation cost of keypoint detection and feature builder as the curvelet transform is based on FFT.

In addition, a set of facial landmarks are used for creating feature vectors in some methods and shape index is widely used to detect landmarks. In [63], keypoints were extracted from a shape dictionary, which was learned on a set of 14 manually placed landmarks on human face. As an extension to [63], [64] used a dictionary of L learned local shapes to detect keypoints, and evaluated them through linear (LDA) and nonlinear (AdaBoost). [65] detected the resolution invariant keypoints and scale-space extreme on shape index images based on scale-space analysis, and used six scale-invariant similarity measures to calculate the matching score. In [66], an entire geometry-based 3D face recognition method was proposed and 17 landmarks were automatically extracted based on the facial geometrical characteristics, which was further extended in [67].

Curve-based: A curve-based method uses a set of curves to construct feature descriptors. It is difficult to define whether it is local feature-based or holistic feature-based, because these curves usually cover the entire face, also they 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 closed curves with different lengths and no intersection. Radial curves are open curves, usually starting from the nose tip.

The level curves can be further divided into equal iso-depth and iso-geodesic curves [13]. The iso-depth curves can be obtained by transposing a plane across the facial surface in one direction and were first introduced by Samir et al. [68]. [69] expanded this work and proposed iso-geodesic curves which are level curves of a surface distance function from the nose tip. However, both of them are sensitive to occlusions, missing parts or larger facial expressions. Thus, radial curves were introduced in [70] and extended in [71]. These curves can better handle the occlusions 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 [72], facial curves in the nose region of a target face were 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, the adaptive regions were selected to match the two identities. In [73], geometric curves from the level sets (circular curves) and streamlines (radial curves) through the Euclidean distance functions of a 3D face 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 Adaboost algorithm [74], which is a well-known machine learning-based feature selection method. Using the selected curves for face recognition, time was reduced from 2.64 seconds to 0.68 seconds. It was proved that the feature selection method could effectively improve system performance. To select high discriminative feature vectors and improve computation efficiency, Angular Radial Signatures (ARSs) was proposed by lei et al. [75]. It was described as a set of curves emitting from the nose tip (as the origin of the facial range images) at intervals of θ radians.

Another type of facial curves was introduced by Berretti et al. [76]. It utilized SIFT to detect keypoints of 3D depth images and connected the keypoints to form the facial curves. A 3D face could be represented by a set of facial curves built by matched keypoints. In [77], There were also 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 [77]. A sequence of RAIKs generated from the surface patch around keypoints can be represented by 2D images so that the certain characteristics of the represented 2D images can have a positive impact on matching accuracy, speed and robustness. [78] introduced Nasal patches and curves. First, seven landmarks on 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 and were filtered by the genetic algorithm to select more stable features against facial expressions. Compared with previous methods, this method has shown high-class separability. Recently, [79] presented a geometry and local shape descriptor based on the Wave Kernel Signature (WKS) [80] to overcome the distortions caused by face expressions.

Local surface-based: One of the representative local feature-based methods is Local Binary Pattern (LBP), introduced by Ojala et al. [81]. It was initially used for 2D images. The local geometric features extracted from some regions of a face surface can be robust to the face expression variations [13]. LBP was used to represent the facial depth and normal information of each face region in [82], where a feature-based 3D face division pattern was proposed to reduce the influence of facial local distortion. Recently, [83] used the LBP algorithm to extract features of a 3D

depth image, and used the SVM algorithm to classify them. The feature extraction time of each depth map in Texas-3D was reduced to 0.1856 seconds while [57] needs 23.54 seconds. Inspired by LBP, [84] proposed the multi-scale and multi-component local normal patterns (MSMC-LNP) descriptor, which can describe normal facial information more compactly and differently. The Mesh-LBP method was used in [85], where LBP descriptors were directly applied on the 3D face mesh surface, fusing both shape and texture information.

Another type of local feature-based methods is based on geometric features. [86] proposed a low-level geometric feature approach, which extracts region-based histogram descriptors from a facial scan. The feature regions include nose and eyes-forehead, which are comparatively less affected by the deformation caused by facial expressions. In this paper, a Support Vector Machine (SVM) and the fusion of these descriptors at both features and score level were applied to improve the accuracy. In [87], a covariance matrix of the feature was used as the descriptor for 3D shape analysis, not the feature itself. Compared with feature-based vectors, covariance-based descriptors can fuse and encode all types of features into a compact representation [88]. Their work was expanded in [88].

There are other local feature-based methods. In [89], local surface descriptors were constructed around keypoints, which were defined by checking the Curvelet coefficient in each subband. 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 [90] used the regional boundary sphere descriptor (RBSR), which reduced the computational cost and improved the classification accuracy. [91] proposed a local derivative mode (LDP) descriptor based on local derivative changes. It can capture more detailed information than LBP. An extension to this work was described in [92]. Recently, Yu et al. [93] recommended utilizing the ICP (Iterative closest point) 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 the accuracy as they are less sensitive to expression or occlusion.

Summary: Most local feature-based 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 one face. The main objective of local feature-based methods is to extract distinctive compact features [13]. Table II summarizes the local feature-based methods, which are further categorized into keypoint-based, curve-based and local surface-based, as recapped below:

- The keypoint-based methods are assorted into three groups: SIFT-based, mesh-based, and landmarks. There are two important points worth noting: the selection of effective keypoints and construction of feature descriptor. If the amount of keypoints is too excessive, the computational cost will increase. However, if the keypoints are too sparse, some key features will be lost and the recognition performance may be affected. In addition, the algorithms for measuring the neighborhood of keypoints are very important as the geometric relationships of keypoints are used to build feature vectors.
- The curved-based methods are broadly classified into level curve-based and radial curve-based methods. Since the level curves are sensitive to the occlusions and missing parts, most curve-based methods use radial curves. Generally, a reference point is required in a curved-based method. The nose region is rigid and contains more distinctive shape features than other regions, so it is used as the reference point in most curve-based methods [13]. Therefore, the detection of nose tip is a crucial step in these methods. Its incorrect position can affect the extraction of curves and the performance of the face recognition system.
- The local surface-based methods are divided into LBP-based, geometric features-based and others. Some of them also need high accuracy of the nose tip detection as the nose tip is used for face segmentation. Most local surface-based methods are robust to facial expressions and postures as the feature vectors are extracted from rigid regions of a face surface.

C. Holistic-based methods

Compared with the local-based methods, holistic-based methods extract features from the entire 3D face surface. They are very effective and can perform well under the complete, frontal, and expression-invariance 3D faces. Common techniques used by holistic-based methods include ICP, Eigenfaces (PCA) and Fisherfaces.

Table III summarizes the remarkable endeavours that have been made in this area. An intrinsic coordinate system for 3D face registration was proposed by Spreeuwers [94]. This system is based on a vertical symmetry plane passing through the nose, 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 new transformation system. [95] presented a method named UR3D-C, which used LDA to train dataset and compress the biometric signature to just 57 coefficients. It still shows a high discriminant under the compact feature vectors. Bounding sphere representation (BSR), introduced in [96], was used to represent both the depth and 3D geometric shape information by projecting the preprocessed 3D point clouds on the bounding spheres. Shape-based Spherical Harmonic Features (SHF) was proposed in [97], where SHFs were calculated based on the spherical depth map (SDM). The SHF can capture the gross shape and fine surface details of a 3D face through the energies contained in the spherical harmonics at different frequencies. [98] used 2DPCA to extract features and employed Euclidean distance for matching. In [99], the authors 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. The pixel in an input scan is verified as a nose tip if the mean square error between a candidate feature vector and its projection on the Eigenface space is less than a predefined threshold.

[100] introduced a Rigid-area Orthogonal Spectral Regression (ROSR) method, where the curvature information is used to segment facial rigid area and OSR is used to extract discriminant feature. In [101], a 3D point cloud is registered in the inherent coordinate system with the nose tip as the origin, and a two-layer ensemble classifier is used for face recognition. A local facial surface descriptor was proposed by [102]. This descriptor is constructed based on three principal curvatures estimated by asymptotic cones. The asymptotic cone is an essential extension of the asymptotic direction to the mesh model. It allows the generation of three principal curvatures representing the geometric characteristics of each vertex. [103] proposed a region-based 3D deformable model (R3DM), which is formed from the densely corresponding faces. Recently, Kernel-based PCA is used for 3D face recognition. Due to the nature of face exhibiting non-linear shapes, non-linear PCA was used in [104] to extract 3D face features as it has a notable benefits to data representation in high-dimensional space.

Based on the discussion above, most holistic-based methods have faster speed and lower computational complexity, but they are not suitable for handling occluded faces or missing part faces. In addition, variations in pose and scale may affect the recognition performance of global features, because the holistic-based algorithms create discriminating features based on all the visible facial shape information. This requires accurate normalization for pose and scale. However, it is not easy to obtain accurate pose normalization under noisy or low-resolution 3D scanning.

TABLE II: Local feature-based techniques.

Author/year	Category	Methods	Advantage	Limitation	Database	RR (Rank-1, %)
Berretti et al. (2011) [50]	SIFT keypoint	Covariance matrix, X^2 dist	Partial facial	Keypoints redundancy	FRGC v2	89.2 (Partial faces)
Li et al. (2011) [58]	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 et al. (2011) [65]	Landmarks	SVM-based fusion, six similarity measures	Simple preprocessing, noise, resolution	Occlusions	FRGC v2	96.2
Inan and Halici (2012) [51]	SIFT keypoint	Cosine dist	Neutral expression	Noise	FRGC v2	97.5
Berretti et al. (2012) [76]	Curve	Sparse	Missing parts	Large pose, expression	FRGC v2 GavabDB UND	95.6 97.13 75
Li and Da (2012) [72]	Curve	PCA	Expression, hair occlusion	Exaggerated expressions	FRGC v2	97.80
Ballihi et al. (2012) [73]	Curve	Euclidean dist, AdaBoost	Efficient, data storage	Occlusions	FRGC v2	98
Berretti et al. (2013) [60]	Mesh-based keypoint	X^2 dist	Missing parts	Low accuracy	UND	77.1
Smeets et al. (2013) [57]	Mesh-based keypoint	Angles comparison	Expression, partial data	Noise	Bosphorus FRGC v2	93.7 89.6
Creusot et al. (2013) [64]	Mesh-based landmark keypoint	Linear (LDA), non-linear (AdaBoost)	Expression	Complexity, occlusions	FRGC v2 Bosphorus	- -
Tang et al. (2013) [82]	Local surface (LBP-based)	LBP, Nearest-neighbor (NN)	Expression	Occlusion, missing data	FRGC v2	94.89
Lei et al. (2013) [86]	Local surface (Geometric feature)	SVM	Expression	Occlusion	FRGC v2 BU-3DFE	95.6 97.7
Elaiwat et al. (2013) [89]	Local surface	Curvelet transform	Illumination, expression	Occlusion	FRGC v2	-
Drira et al. (2013) [71]	Curve	Riemannian framework	Pose, missing data	Extreme expression, complexity	FRGC v2	97.7
Li et al. (2014) [84]	Local surface (LBP-based)	ICP, Sparse-based	Expression, fast	Pose, occlusion	FRGC v2	96.3
Berretti et al. (2014) [61]	Mesh-based keypoint	Classifier	Occlusions, missing parts	Noise, low-resolution image	Bosphorus	94.5
Lei et al. (2014) [75]	Curve	KPCA, SVM	Efficient, expression	Occlusion	FRGC v2 SHREC08	- -

TABLE II: Local feature-based techniques (continued).

Tabia et al. (2014) [87]	Local surface (Geometric features)	Riemannian metric	Expression	Occlusion	GavabDB	94.91
Vezzetti et al (2014) [66]	Landmark keypoint	Euclidean distance	Expression, occlusion	Low accuracy	Bosphorus	-
Li et al. (2015) [59]	Mesh-based keypoint	Gaussian filters, fine-grained matcher	Expression, occlusion, registration-free	Cost	Bosphorus	96.56
Elaiwat et al. (2015) [62]	Mesh-based keypoint	Curvelet transform, cosine dist	Illumination, expressions	Occlusion	FRGC v2	97.1
Al-Osaimi (2015) [77]	Curve	Euclidean dist	Fast, expression	Occlusion	FRGC	97.78
Ming (2015) [90]	Local surface	Regional, global regression	Large pose, efficient	Patches detection	FRGC v2 CASIA BU-3DFE	- - -
Guo et al. (2016) [53]	keypoint	Rotational Projection Statistics (RoPS), average dist	Occlusion, expression and pose	Cost	FRGC v2	97
Soltanpour and Wu (2016) [52]	SIFT keypoint	Histogram matching	Expression	Pose	FRGC v2	96.9
Lei et al. (2016) [54]	SIFT keypoint	Two-Phase Weighted	Missing parts, occlusions, data corruptions	Extreme pose, expression	FRGC v2	96.3
Emambakhsh and Evans (2016) [78]	Curve	Mahalanobis, cosine dist	Expression, single sample	Occlusion	FRGC v2	97.9
Werghi et al. (2016) [85]	Local surface (LBP-based)	Cosine, X^2 dist	Expression, missing data	pose	BU-3DFE Bosphorus	- -
Hariri et al. (2016) [88]	Local surface (Geometric features)	Geodesic dist	Expression, pose	Partial occlusions	FRGC v2	99.2
Soltanpour et al. (2017) [91]	Local surface (LDP)	ICP	Expression	Extreme pose, missing data	FRGC v2 Bosphorus	98.1 97.3
Deng et al. (2017) [55]	SIFT keypoints	Riemannian kernel sparse coding	Low-complex	Expression, occlusion	FRGC v2	97.3
Abbad et al. (2018) [79]	Curve	Angles comparison	Expression, time consumption	Occlusions and missing data	GavabDB	99.18
Soltanpour et al. (2019) [92]	Local surface (LDP)	ICP	Computational cost	Pose	FRGC v2	99.3
Shi et al. (2020) [83]	Local surface (LBP-based)	LBP, SVM	Low consumption	Pose, occlusions	Texas-3D	96.83

D. Hybrid methods

Hybrid 3D face recognition methods combine different types of approaches (local-based and holistic-based) and apply both local and global features for face matching. They can handle more face variances such as expression, pose, and occlusion via combining different feature extraction techniques. Recent hybrid methods are compared in Table IV.

[105] used an automatic landmark detector to estimate poses and detect occluded areas, and used facial symmetry to deal with missing data. [106] proposed a hybrid matching scheme using multiscale extended LBP and SIFT-based strategy. In [107], the problem of external occlusions 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 are used to compute the masked projection. This method relies on the accurate nose tip detection. The performance is adversely affected if the data has some occlusions covering the nose area. [108] extended this work in 2013. [109] proposed a scale-space based representation for 3D shape matching which is stable against surface noise.

In [110], 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 occlusions. Later, they improved the registration method and proposed an across-pose method in [111]. [112] 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 face recognition systems use both local features and global features. Thus, their structures may be more complicated than those of the local-based or holistic-based methods. The hybrid approaches could achieve better recognition accuracy at higher computational cost. In addition,

TABLE III: Holistic-based techniques.

Author/year	Methods	Advantage	Limitation	Database	RR (Rank-1, %)
Spreeuwers (2011) [94]	PCA-LDA	Less registration time	Expression, occlusions	FRGC v2	99
Ocegued et al. (2011) [95]	Ll norm, ICP, LDA, Simulated Annealing algorithm	Speed efficient	Expression, occlusions	FRGC v2	99.7
Ming and Qiuqi (2012) [96]	Robust group sparse regression model (RGSRM)	Expression, pose	Distorted images	FRGC v2 CASIA	- -
Peijiang et al. (2012) [97]	-	Faster, cost-effective	Expression, occlusion	SHREC2007 FRGC v2 Bosphorus	97.86 96.94 95.63
Taghizadegan et al. (2012) [98]	PCA, Euclidean distance	Expression	Occlusion	CASIA	98
Mohammadzade and Hatzinakos (2012) [99]	PCA	Computation, expression	Occlusion, pose	FRGC	-
Ming (2014) [100]	PCA, Spectral Regression, the orthogonal constraint	Expression, computational cost, storage space	Occlusions	FRGC v2	95.24
Ratyal et al. (2015) [101]	PCA, Mahalanobis Cosine (MahCos)	Pose, expression	Occlusion, missing part	GavabDB FRGC v2	100 98.93
Tang et al. (2015) [102]	principal curvatures	Computational cost	Expression, occlusion	FRGC v2	93.16
Gilani et al. (2017) [103]	PCA, use CNN for landmark detection	Occlusions	Faster, expressions, poses	Bosphorus	98.1
Peter et al. (2019) [104]	Kernel-based PCA	Higher accuracy rate	-	FRGC v2	-

TABLE IV: Hybrid techniques.

Author/year	Methods	Advantage	Limitation	Database	RR (Rank-1, %)
Passalis et al. (2011) [105]	PCA	Pose, occlusion, missing data	Expression, low accuracy	UND	-
Zhang et al. (2012) [106]	SIFT-based, extended LBP	Registration-free (frontal)	Large pose (alignment required)	FRGC v2	97.6
Alyuz et al. (2012) [107]	ICP, PCA, LDA	Occlusions	Expression	Bosphorus	83.99
Fadaifard et al. (2013) [109]	L1-norm	Noise, computational efficiency	Occlusions, expression	GavabDB	86.89
Alyuz et al. (2013) [108]	ICP, PCA, LDA	Occlusions, missing data	Expression	Bosphorus UMBDB	- -
Bagchi et al. (2014) [110]	ICP, PCA	Pose, occlusions	Pose	Bosphorus	91.3
Bagchi et al. (2015) [111]	ICP, KPCA	Pose	Expression	GavabDB Bosphorus FRAV3D	96.92 96.25 92.25
Liang et al. (2017) [112]	HK classification	Pose	Expression	Bosphorus	94.79

similar to holistic-based methods, its face registration is a very important step, especially for overcoming the pose-variance and occlusion-variance.

IV. Deep learning-based 3D face recognition

In the last decade, deep neural networks have become one of the most popular techniques for face recognition. Compared with the conventional ones, deep learning-based methods have great advantages over image processing [113]. For conventional methods, the key step is to find robust feature points and descriptors based on geometric information of 3D face data [114]. Compared with the end-to-end deep learning models, these methods have good recognition performance, but involve relatively complex algorithmic operations to detect key features [114]. While for deep learning-based methods, robust face representations can be learned by training a deep deep neural networks on large datasets [114], which can hugely improve the face recognition speed.

There are a variety of deep neural networks for facial recognition and convolutional neural networks (CNN) are the most popular ones. A CNN usually consists of convolutional layers, pooling layers, and fully-connected (FC) layers. The purpose of a convolutional layer is to extract features from the input data. Each convolutional layer performs convolution operation with a filter kernel and applies a nonlinear transfer function. The objective of the pooling layers is to reduce the dimensions of the feature maps by integrating the outputs of neuron clusters of one layer into a single neuron in the next layer [14]. The robust and discriminative feature representation learned via CNN can significantly improve the performance of face recognition. Fig. 4 depicts a common face recognition process based on Deep-CNN (DCNN). During the training phase, a set of training data is preprocessed first, such as alignment, resizing, etc., to

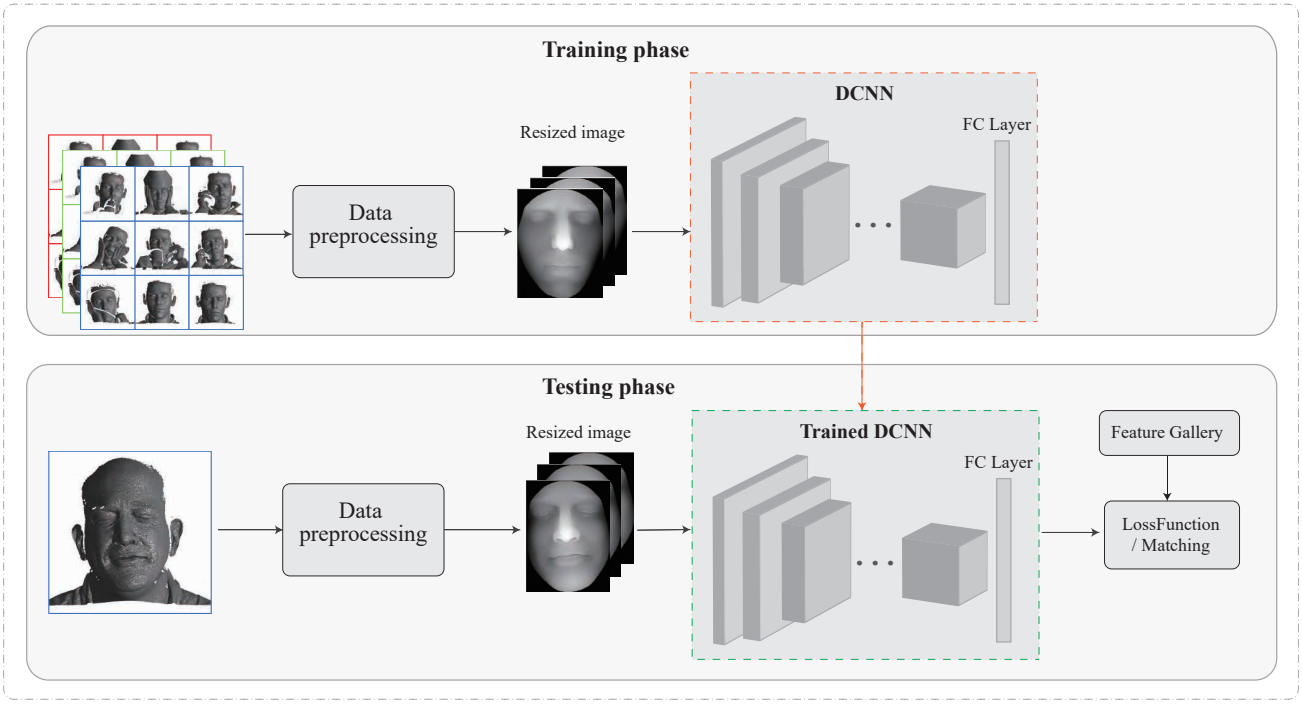


Fig. 4: An overview of 3D deep learning-based face recognition methods.

generate a unified feature map and fit the input tensor of the DCNN architecture (e.g. the height, width, and channel of the feature map and the number of images). Then, the DCNN is trained by the preprocessed maps. In the testing phase, the feature representation of a probe is obtained from the trained DCNN and used to match with features in a given gallery. Before discussing the 3D face recognition methods using DCNN, we have a quick review on some typical 2D deep learning-based methods as their networks are still being used by some 3D methods.

A. 2D Face recognition

With the application of CNN, the performance of 2D face recognition systems ([115], [116], [117], [16], [118], [17]) has been significantly improved. In these systems, face representations are directly learned from 2D facial images by training deep neural networks on large datasets. DeepFace [115] model is a nine-layer deep neural network which is trained on a labeled dataset including 4M facial images with over 4k identities. A 3D model-based alignment method is used and 97.35% accuracy is achieved on the LFW [4] dataset. The authors later extended this work in [119] and believed that the performance of CNN may reach a saturation point when the size of the training dataset increases.

The DeepId series methods (DeepID [116], DeepID2 [120], DeepID2+ [121], DeepID3 [122]) extract deep features from various face regions. They incrementally improve the performance and reduce the error rate on the LFW dataset. [117] proposed an inception DCNN architecture and created a 22-layers deep network named GoogLeNet model by repeating the Inception layers. The GoogLeNet model was trained on a large dataset with 200M face images of 8M identities. It utilizes metric learning algorithms and directly learns a mapping from face images to the compact Euclidean space. The embedding itself is optimized by a triplet loss during the network training. This is followed by VGG-Face [16], which fine-tunes their model by a triplet-based metric learning method like FaceNet [118]. They also provided a large face dataset with 2.6M 2D images from 2,622 identities. Consequently, ResNet [17] proposed a residual learning framework to simplify network training and evaluated the residual nets of up to 152 layers on the ImageNet dataset [123]. Its depth is 8 times that of the VGG-Face [16], but still shows lower complexity.

B. 3D Face recognition

As discussed above, deep learning-based 2D face recognition has made great achievements, and its performance is extremely high, almost close to 100% on some specific databases (such as LFW). The high recognition rate of 2D face recognition proves 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 pose, illumination and expression, more researchers are attracted to apply DCNN for 3D face recognition. Table V enlists the latest 3D face recognition techniques based on DCNN, and table VI compares the recognition rate on different databases.

VGG-based: Kim et al. [114] proposed the first 3D face recognition model with DCNN. Their CNN architecture uses the VGG-Face [16] trained on 2D face images and then fine-tunes the CNN with augmented 2D depth maps. The last

FC layer of the VGG-Face is replaced with a new last FC layer and a softmax layer. At the new last layer, weights are randomly initialized from a Gaussian distribution, with a mean of zero and a standard deviation of 0.01 [114]. In addition, 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. In [114], a multi-linear 3D Deformable Model (3DMM) is used to generate more expressions, including variations in both shape (α) and expression (β). A 3D point cloud can then be represented by[114]:

$$\mathbf{X} = \overline{\mathbf{X}} + \mathbf{P}_s\alpha + \mathbf{P}_e\beta \quad (1)$$

where $\overline{\mathbf{X}}$ is the average facial point cloud, $\mathbf{P}_s$ is the shape information provided by the Basel Face Model [124], and $\mathbf{P}_e$ is the expression provided by FaceWarehouse [125]. The expression variations are created by randomly changing the values of expression β parameters in the 3DMM. The randomly generated rigid transformation matrices are applied to an input 3D point cloud to demonstrate the pose variances [114]. At data preprocessing stage, a nose tip is first found in a 3D point cloud, then the 3D point cloud is cropped within a 100 mm radius. The rigid feature between the 3D face model and the reference face model is used to align the face 3D model. In order to fit the input size of its CNN architecture, the aligned 3D scan is orthogonally projected onto 2D image to generate a $224 \times 224 \times 3$ depth map. In addition, patches are randomly removed from the depth map to simulate hard occlusion. The model was evaluated on three public 3D databases: Bosphorus [30], BU3D-FE [26] and 3D-TEC [35], and the recognition rates were 99.2%, 95.0% and 94.8%, respectively.

Deep 3D Face Recognition Network (FR3DNet) [42] was trained on 3.1M 3D faces and specifically designed for 3D face recognition. It is also based on VGG face [16]. A rectifier layer is added for every convolutional layer. Compared with Kim's work [114], a much larger dataset is generated and expanded by new identities. A new face $\hat{\mathbf{F}}$ is generated from a pair of faces ($\mathbf{F}_i$, $\mathbf{F}_j$) with the maximum non-rigid shape difference [42]:

$$\hat{\mathbf{F}} = \frac{\mathbf{F}_i + \mathbf{F}_j}{2} \quad (2)$$

These synthetic faces generated by this method have richer shape changes and details than the statistical face models [124]. However, the computational cost is very high as they are all generated from high dimensional raw 3D faces. In addition, 15 synthetic cameras were deployed at the front hemisphere of the 3D face to simulate pose variations and occlusions in each 3D scan. To fit the input of FR3DNet, 3D point cloud data is preprocessed to a $160 \times 160 \times 3$ image [124]. 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 (x, y grid) algorithm [126]. They are the depth map z(x, y), azimuth map θ(x, y) and elevation map ϕ(x, y), where θ and ϕ are the azimuth and elevation angles of the normal vectors of 3D point cloud surface. The experiments were conducted on most public databases and the highest recognition accuracy was achieved on the BU3D-FE [26] database, reaching 98.64%.

ResNet-based: Y. Cai, Y. Lei and M. Yang et al. [5] designed three deep residual networks with different layers based on the ResNet [17], named as Pre-ResNet-14, Pre-ResNet-24 and Pre-ResNet-34. It is worth mentioning that a multi-scale triplet loss supervision is constructed by combining a softmax loss and the two triplet loss supervision on the last fully connected layer and the last feature layer [5]. To enlarge the size of the training set, the data was augmented in three ways: pose augmentation based on 3D scan, resolution and transformational augmentation based on range images [5]. For the preprocessing algorithm, raw 3D data is 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 only the nose tip) and one entire facial region are generated from the raw 3D data [5]. The most important part of this method is detecting the nose tip and two pupils. The 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 9 (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 could 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 so that the computational cost can be reduced. 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: the FRGC v2, Bosphorus, BU-3DFE, and 3D-TEC datasets with 100%, 99.75%, 99.88%, and 99.07%, respectively.

Lin at al. [127] also adopted the ResNet-18 as their network. The big difference to other work is their data augmentation method. Instead of reconstructing 3D face samples from other raw data, they generated the the feature tensors directly based on the 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 $\mathbf{F} = [f_i, ...f_{13}]$ and the sub feature is $\mathbf{SubF}_i$. 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 [127]:

$$\mathbf{F^k} = U_{i=1}^{13} SubF_i^k, k = 1, .., K \quad (3)$$

where K is the number of 3D face samples of the same person. A new feature set could be synthesized by randomly choosing the i^{th} sub feature set from the K samples. The network achieved very competitive performance on both Bosphrous and BU3D-FE databases with accuracy of 99.71% and 96.2% , respectively.

Tan at al. [128] designed a framework to specifically process the low-quality 3D data being captured by portable 3D acquisition techniques like 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, noises and a random pose variation are added to the face scan. Then the new point cloud is projected onto a 2D plane with 1000 grids of the same size and a sparse face of 1,000 points is obtained by randomly selecting a point from each grid. Six sparse faces are generated from each face scan and passed to the DRNet to generate a new dense point cloud. Then the fused data is used as the input to Face Recognition Network (FRNet) which is also based on ResNet-18. Compared with FR3DNet, its facial recognition rate on UMBDB is higher, reaching 99.2%.

Others: Feng et al [129] adopted two DCNNs for feature extraction: one for color image, and the other for depth map built from 3D raw data. The output of the two feature layers was fused as the final input to an artificial neural network (ANN) recognition system. This ANN recognition system was tested on CASIA (V1) to compare the recognition rates by separately using the 2D feature layer, 3D feature layer, and the fusion of 2D and 3D features layers. A higher RR (98.44%) was obtained with the fusion features.

Olivetti at al. [130] 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 a low amount of parameters to tune. All their work was based on the Bosphorus database, which only contains 105 identities with 4,666 images. In order to obtain enough training samples, they augmented the data by rotating the original depth map (clockwise 25, counterclockwise 40) and creating a horizontal mirror for 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 proven in [131]. The input data is a three-channel image including the 3D facial depth map, the shape index and the curvedness, which can enhance the accuracy of the network. A 97.56% recognition rate was achieved on the Bosphorus database.

Xu at al. [132] designed a dual neural network to reduce the impact of the number of training samples. The network consists of a dual-channel input layer that can fuse the 2D texture image and 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 the similarity. The gray-scale depth map obtained from the point cloud, combining with the corresponding 2D texture, is used as the dual-channel input. The most important step of the preprocessing algorithm is the face hole filling, which provides a better intact face. The basic idea is to 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.

Zhang Z, Da F, Yu Y. [133] proposed a network structure similar to PointNet++. It contains three set abstraction (SA) modules: curvature-aware point sampler, neighbors grouper and multi-layer perceptrons (MLP). The first two SA are used to extract local features and the last one is used to aggregate global feature. The most important part of this work is the training set. All the training sets are unreal data, that are synthesized by sampling from a statistical 3D Morphable Model of face shape and expression based on the GPMM [134]. This method addresses the problem of lacking a large training dataset. After classification training, triplet loss is used to fine-tune the network with real faces, which can get better results.

Mu at al. [135] proposed a lightweight CNN for 3D face recognition, especially for low-quality data. This network contains 4 blocks which have 32, 64, 128, and 256 convolution filters, respectively. The feature maps from these four convolutional blocks are captured by different Receptive Fields (RFs), down sampled to a fixed size by max-pooling, and integrated to form another conversion block. This process is completed by the Multi-Scale Feature Fusion (MSFF) module. The aim is to efficiently improve the representation of low-quality face data. The Spatial Attention Vectorization (SAV) module is used to replace the Global Average Pooling (GAP) 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 [135]. In addition, three methods are used to augment the training data: pose generating (by adjusting virtual camera parameters), shape jittering (by adding Gaussian noise to simulate rough surface changes), and shape scaling (by zooming in the depth face image with 1.1 times). At data preprocessing stage, similar to the above methods, a 10×10 patch surface is first cropped around a given nose tip with outliers removal. Then the cropped 3D point cloud is projected into a 2D space (depth surface) to generate a normal map image.

Bhople at al. [136] proposed a network based on the PointNet architecture. It directly uses point cloud as input and uses the Siamese network for similarity learning. In addition, a way to augment database at the point cloud level is provided.

Cao at al. [137] believed that the key to a reliable face recognition system is rich data sources. They used ANN as the network architecture, but paid more attention to data acquisition. A holoscopic 3D (H3D) face image database was created, which contains 154 raw H3D images. H3D imaging is recorded by using a regularly arranged array of small lenses, which are closely packed together and connected with a recording device [137]. Therefore, it can display 3D images with continuous parallax and full-color images can be viewed

TABLE V: Deep learning-based techniques.

Author/year	Network	Layers	Input	Matching	Database	RR (%)(Rank-1)
Kim et al. (2017) [114]	Finetuning VGG	16 convolutional, 3 FC, 1 softmax	$224 \times 224 \times 3$	Cosine distance	Bosphorus	99.2
Zulqarnain et al. (2018) [42]	FR3DNet	13 convolutional, 3 FC, 1 softmax	$160 \times 160 \times 3$	Cosine distance	Texas-3D	100.0
Feng et al. (2019) [129]	ANN	2 DCNN	Depth map	-	CASIA	98.44
Cai et al. (2019) [5]	Pre-ResNet-34, Pre-ResNet-24, Pre-ResNet-14	-	$96 \times 96 \times 3$	Euclidean distance	FRGC v2	100
Lin at al. (2019) [127]	ResNet-18	17 convolutional, 1 FC	$256 \times 256 \times 3$	Similarity tensor calculated from 2 feature tensors	Bosphorus	99.71
Olivetti at al. (2019) [130]	MobileNetV2	-	$224 \times 224 \times 3$	-	Bosphorus	97.56
Tan at al. (2019) [128]	ResNet-18	17 convolutional, 1 FC	$256 \times 256 \times 3$	Cosine distance	CASIA	99.7
Xu at al. (2019) [132]	LeNet5	Two parallel CNNs (4 convolutions, 4 pooling, and 1 FC)	RGB image with depth map	Euclidean distance	CASIA	-
Zhang at al. (2019) [133]	Data-Free Point Cloud Net (similar to PointNet++)	3 set abstraction (SA) modules	3D Point Cloud	Cosine similarity	FRGC v2	98.73
Mu at al. (2019) [135]	-	4 convolution blocks, a MSFF module and a SAV module	Low quality input	Cosine distance	Lock3DFace	81.02
Cao at al. (2020) [137]	ANN	-	H3D image	-	H3D [137]	-
Dutta at al. (2020) [138]	SpPCANet	Convolution layer, nonlinear processing layer, feature-pooling layer	Depth images	linear SVM [139]	Frav3D	96.93

in a wider viewing area. Wavelet transform is used for feature extraction, as it performs well in the presence of illumination changes and face orientation changes, also reduces image information redundancy and retains the most important facial features [137]. This is definitely a new direction for 3D face recognition, but the accuracy of this method is quite low, only reaching slightly higher than 80%.

Dutta at al. [138] proposed a lightweight sparse principal component analysis network (SpPCANet). It includes three parts: convolutional layer, nonlinear processing layer and feature merging layer. For data preprocessing, common ways are used to detect and crop the face area. First an ICP-based registration technology is used to register a 3D point cloud, then the 3D point cloud is converted into a depth image and finally all faces are cropped into rectangles based on the position of the nose tip. The system obtained 98.54% recognition rate on Bosphorus3D.

Summary: In this section, we review the deep learning-based 2D and 3D face recognition techniques. Most of the DCNN-based 3D methods achieve very high recognition accuracy and run at fast speed. For example, [5] only required 0.84 seconds to identify a target face from a gallery of 466 faces. [42] got 100% recognition rate on Texas-3D. There are three important parts in a DCNN-based system: training set, data preprocessing and network architecture. The deep learning-based methods always require a large number of datasets to train the network. The lack of large-scale 3D face datasets is still an open problem in DCNN-based 3D face recognition research. Before passing the data to the network, appropriate data preprocessing can improve accuracy, as CNN usually has less tolerance for pose changes. In addition, adopting a suitable DCNN is also important. In the above reviewed works, most of them use a single CNN but few use dual CNN such as [132]. The reorganization of CNNs may also be a topic of future research.

V. Discussion

In the past decade, 3D face recognition has achieved significant growth in 3D face databases, recognition rates, and robustness to face data variance, such as low-resolution, expression, pose and occlusion. The conventional methods and deep learning-based methods are thoroughly reviewed in Section 3 and Section 4, respectively. Based on the feature extraction algorithms, the conventional methods are divided into three types: local feature-based, holistic-based 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 the detected key-points. Compared to the global descriptors for the holistic-based methods, the number of extracted local descriptors is related to the content of input face (entire

TABLE VI: RR (%)(Rank-1) of DCNN-based methods on other databases. HQ: high quality image. LQ: low quality image.

Reference	FRGC v2	BU3D-FE	BU4D-FE	Bosphorus	CASIA	GavabDB	Texas-3D	3D-TEC	UMBDB	ND-2006
Kim et al. [114]	-	95.0	-	99.2	-	-	-	94.8	-	-
FR3DNet[42]	97.06	98.64	95.53	96.18	98.37	96.39	100.00	97.90	91.17	95.62
FR3DNetFT. [42]	99.88	99.96	98.04	100.0	99.74	99.70	100.0	99.12	97.20	99.13
Feng et al. [129]	-	-	-	-	85.93	-	-	-	-	-
Cai et al. [5]	100	99.88	-	99.75	-	-	-	99.07	-	-
Lin at al. [127]	-	96.2	-	99.71	-	-	-	-	-	-
Olivetti at al. [130]	-	-	-	97.56	-	-	-	-	-	-
Tan at al. [128]	-	-	-	99.2	99.7	-	-	-	99.2	-
Zhang at al.[133]	92.74	-	-	93.38	-	-	-	-	-	-
Zhang at al. (ft)[133]	98.73	-	-	97.5	-	-	-	-	-	-
Mu at al. [135]	-	-	-	91.27 (HQ) 90.70 (LQ)	-	-	-	-	-	-
Dutta at al. (2020) [138]	-	-	-	98.54	88.80	-	-	-	-	-

or partial). It is commonly assumed that only a small number of facial regions are affected by occlusion, partial missing or distortion caused by data corruption, while most other regions persist unchanged. A face representation is derived from the combination of many local descriptors. Therefore, local facial descriptors are not compromised when dealing with the deformation of a few parts caused by facial expressions or occlusions [54].

- A global representation is extracted from an entire 3D face, which usually makes the holistic-based methods compact and therefore computationally efficient. In addition, these methods can achieve great accuracy in the presence of complete neutral faces. However, they rely on the availability of full face scans and are sensitive to face alignment, occlusion, and data corruption. Therefore, face registration is a very important step for holistic-based methods.
- The hybrid methods combine the algorithms of extracting local features and global features. They can handle more conditions, such as the pose variance and occlusion variance.

Since 2016, research on DCNN-based 3D face recognition has been carried out. Table VI summarizes the recognition rate of our surveyed methods on different databases under rank-1. Compared with the conventional face recognition algorithms, DCNN-based methods have the advantages of simpler pipelines and higher performance. In general, the deep learning-based methods do not have to perform keypoint detection, face segmentation or feature fusions. Instead, they only need to convert 3D data into a suitable network input format (e.g. 2D images). Moreover, since the pretrained networks are often fine-tuned using the training data generated from 3D faces, the chance of delivering promising performance has been greatly improved. However, they are more reliant on the training set than the conventional methods. Therefore, data augmentation is one of the key challenges we are facing. Besides of the network structure design, data preprocessing also has a huge influence on the recognition performance.

To improve the accuracy and performance of face recognition systems, we discuss the following (future) directions by considering new face data generation, data preprocessing and DCNN design.

- Augmenting new face data. In Section 4, almost every proposed method provides a strategy for augmenting new face data. It is because a large amount of training data is required to train networks. A network trained with sufficient data can better distinguish features, while a small number of samples may cause overfitting. To synthesize new face data, one way is to use 3D deformable facial Model to generate new shape and expression (such as [114], [133]). Another method is to randomly select sub-feature sets from different samples of a person and combine them to generate a new face. Some other methods augment the pose variance by rotating the original data (e.g. [130]).
- Data Preprocessing. It is also a key point of improving face recognition accuracy. A well-known problem of the rigid ICP registration is that it cannot guarantee an optimal convergence [114]. This means it may not be possible to accurately register all 3D faces with different poses to the reference face. Furthermore, CNNs may not have much tolerance to pose deviation. Better conversion techniques (e.g. from 3D faces to 2D images) would also improve face recognition performance.
- Applying appropriate loss functions. There are many CNNs available for 3D face recognition, such as VGG, ResNet, mobileNet and PointNet++. Most researchers adopt one of them as their network. Usually, the network is restructured by changing the FC layer and adding softmax. Recently, applying loss functions to supervise the network layers also has become one active research topic. Using effective loss functions can reduce the

complexity of training and improve feature learning capabilities. For example, [133] adopted multiple loss functions to improve the extraction efficiency.

- Creating large-scale 3D face database. Current 3D face databases are often smaller compared to the counterparts in 2D color face recognition, and nearly all the deep learning-based 3D face recognition methods fine-tune the pre-trained networks on the converted data from 3D faces. Hence, large-scale 3D face databases could enable training from scratch and improve the recognition difficulty, making it more closer to real-world applications.

Besides of the above aspects, researchers can consider combining conventional methods with CNN. For example, the keypoint detection techniques in conventional 3D face recognition methods can be incorporated into the deep learning-based methods to better attend area of interest. 3D face recognition methods for low-quality data (such as low resolution) also need more attention.

VI. Conclusion

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 the 3D face recognition methods proposed in the past decade. The recognition methods are divided into two categories based on their feature extraction methods: conventional and deep learning-based. The conventional techniques are further classified into local-based, holistic-based and hybrid methods. We reviewed these methods by comparing their performance on different databases, the computational cost, and the robust against the expressions, occlusions and pose variations. From the above literature review, we found that local methods can better handle face expressions and occluded images at the cost of higher computation comparing with holistic-based methods. Hybrid methods combined the local and global features can achieve better performance and address the challenges including pose variations, illumination changes and facial expressions, etc.

We reviewed recent advances in 3D face recognition based on deep learning, mainly focusing on face augmentation, data preprocessing and network architectures. CNN is one of the most popular deep neural networks for 3D face recognition. According to the network adopted, the deep learning-based 3D face recognition methods are broadly divided into VGG-based, ResNet-based and others. With these powerful networks, the performance of 3D face recognition has been greatly improved.

We also discussed the involved characteristics and challenges, and provided potential future directions for 3D face recognition. For instance, large-scale 3D face databases are in great need to advance 3D face recognition in the future. We believe our survey will provide comprehensive information to readers and inspire insights in the community.

References

[1] H. Patil, A. Kothari, and K. Bhurchandi, "3-d face recognition: features, databases, algorithms and challenges," *Artificial Intelligence Review*, vol. 44, no. 3, pp. 393–441, 2015.

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

[3] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition," *Computer vision and image understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[4] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," 2008.

[5] Y. Cai, Y. Lei, M. Yang, Z. You, and S. Shan, "A fast and robust 3d face recognition approach based on deeply learned face representation," *Neurocomputing*, vol. 363, pp. 375–397, 2019.

[6] S. Zhou and S. Xiao, "3d face recognition: a survey," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, p. 35, 2018.

[7] D. M. Blackburn, M. Bone, and P. J. Phillips, "Face recognition vendor test 2000: evaluation report," DEFENSE ADVANCED RESEARCH PROJECTS AGENCY ARLINGTON VA, Tech. Rep., 2001.

[8] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002: Evaluation report," Tech. Rep., 2003.

[9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 947–954.

[10] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "Frvt 2006 and ice 2006 large-scale experimental results," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 5, pp. 831–846, 2009.

[11] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," *Pattern recognition letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[12] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens, "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, pp. 710–727, 2011.

[13] S. Soltanpour, B. Boufama, and Q. J. Wu, "A survey of local feature methods for 3d face recognition," *Pattern Recognition*, vol. 72, pp. 391–406, 2017.

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

[15] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*. IEEE, 2018, pp. 471–478.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[19] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," 2018.

[20] A. Colombo, C. Cusano, and R. Schettini, "Umb-db: A database of partially occluded 3d faces," in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2011, pp. 2113–2119.

[21] C. Beumier and M. Acheroy, "Automatic 3d face authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 315–321, 2000.

[22] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.*, vol. 2. IEEE, 2003, pp. 201–204.

[23] A. Moreno, "Gavabdb: a 3d face database," in *Proc. 2nd COST275 Workshop on Biometrics on the Internet, 2004*, 2004, pp. 75–80.

[24] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2d+ 3d face biometrics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 4, pp. 619–624, 2005.

[25] Y. Wang, G. Pan, Z. Wu, and Y. Wang, "Exploring facial expression effects in 3d face recognition using partial icp," in *Asian Conference on Computer Vision*. Springer, 2006, pp. 581–590.

[26] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *7th international conference on automatic face and gesture recognition (FGR06)*. IEEE, 2006, pp. 211–216.

[27] C. Xu, T. Tan, S. Li, Y. Wang, and C. Zhong, "Learning effective intrinsic features to boost 3d-based face recognition," in *European Conference on Computer Vision*. Springer, 2006, pp. 416–427.

[28] C. Conde, A. Serrano, and E. Cabello, "Multimodal 2d, 2.5 d & 3d face verification," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 2061–2064.

[29] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using a multi-instance enrollment representation to improve 3d face recognition," in *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2007, pp. 1–6.

[30] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *European workshop on biometrics and identity management*. Springer, 2008, pp. 47–56.

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

[32] F. B. Ter Haar, M. Daoudi, and R. C. Veltkamp, "Shape retrieval contest 2008: 3d face scans," in *2008 IEEE International Conference on Shape Modeling and Applications*. IEEE, 2008, pp. 225–226.

[33] Y. Baocai, S. Yanfeng, W. Chengzhang, and G. Yun, "Bjut-3d large scale 3d face database and information processing," *Journal of Computer Research and Development*, vol. 6, no. 020, p. 4, 2009.

[34] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3d face recognition database," in *2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*. IEEE, 2010, pp. 97–100.

[35] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Twins 3d face recognition challenge," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–7.

[36] R. C. Veltkamp, S. van Jole, H. Drira, B. B. Amor, M. Daoudi, H. Li, L. Chen, P. Claes, D. Smeets, J. Hermans *et al.*, "Shrec'11 track: 3d face models retrieval." in *3DOR*, 2011, pp. 89–95.

[37] Y. ZHANG, Z. GUO, Z. LIN, H. ZHANG, and C. ZHANG, "The npu multi-case chinese 3d face database and information processing," *Chinese Journal of Electronics*, vol. 21, no. 2, pp. 283–286, 2012.

[38] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, and P. Liu, "A high-resolution spontaneous 3d dynamic facial expression database," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, pp. 1–6.

[39] R. Min, N. Kose, and J.-L. Dugelay, "Kinectfacedb: A kinect database for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1534–1548, 2014.

[40] J. Zhang, D. Huang, Y. Wang, and J. Sun, "Lock3dface: A large-scale database of low-cost kinect 3d faces," in *2016 International Conference on Biometrics (ICB)*. IEEE, 2016, pp. 1–8.

[41] P. Urbanová, Z. Ferková, M. Jandová, M. Jurda, D. Černỳ, and J. Sochor, "Introducing the fidentis 3d face database," *AnthropologicAl review*, vol. 81, no. 2, pp. 202–223, 2018.

[42] S. Zulqarnain Gilani and A. Mian, "Learning from millions of 3d scans for large-scale 3d face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1896–1905.

[43] S. Jia, X. Li, C. Hu, G. Guo, and Z. Xu, "3d face anti-spoofing with factorized bilinear coding," *arXiv preprint arXiv:2005.06514*, 2020.

[44] Y. Ye, Z. Song, J. Guo, and Y. Qiao, "Siat-3dfe: A high-resolution 3d facial expression dataset," *IEEE Access*, vol. 8, pp. 48 205–48 211, 2020.

[45] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao, "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*, 2020, pp. 601–610.

[46] S. Z. Gilani and A. Mian, "Towards large-scale 3d face recognition," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2016, pp. 1–8.

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

[48] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[50] S. Berretti, A. Del Bimbo, and P. Pala, "3d partial face matching using local shape descriptors," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, 2011, pp. 65–71.

[51] T. Inan and U. Halici, "3-d face recognition with local shape descriptors," *IEEE transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 577–587, 2012.

[52] S. Soltanpour and Q. J. Wu, "Multimodal 2d–3d face recognition using local descriptors: pyramidal shape map and structural context," *IET Biometrics*, vol. 6, no. 1, pp. 27–35, 2016.

[53] Y. Guo, Y. Lei, L. Liu, Y. Wang, M. Bennamoun, and F. Sohel, "Ei3d: Expression-invariant 3d face recognition based on feature and shape matching," *Pattern Recognition Letters*, vol. 83, pp. 403–412, 2016.

[54] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, "A two-phase weighted collaborative representation for 3d partial face recognition with single sample," *Pattern Recognition*, vol. 52, pp. 218–237, 2016.

[55] X. Deng, F. Da, and H. Shao, "Efficient 3d face recognition using local covariance descriptor and riemannian kernel sparse coding," *Computers & Electrical Engineering*, vol. 62, pp. 81–91, 2017.

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

[57] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "meshsift: Local surface features for 3d face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158–169, 2013.

[58] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen, "Expression robust 3d face recognition via mesh-based histograms of multiple order surface differential quantities," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 3053–3056.

[59] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "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, pp. 128–142, 2015.

[60] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Matching 3d face scans using interest points and local histogram descriptors," *Computers & Graphics*, vol. 37, no. 5, pp. 509–525, 2013.

[61] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3d face scans," *The Visual Computer*, vol. 30, no. 11, pp. 1275–1292, 2014.

[62] S. Elaiwat, M. Bennamoun, F. Boussaïd, and A. El-Sallam, "A curvelet-based approach for textured 3d face recognition," *Pattern Recognition*, vol. 48, no. 4, pp. 1235–1246, 2015.

[63] C. Creusot, N. Pears, and J. Austin, "Automatic keypoint detection on 3d faces using a dictionary of local shapes," in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*. IEEE, 2011, pp. 204–211.

[64] Creusot, Clement, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3d meshes," *International journal of computer vision*, vol. 102, no. 1-3, pp. 146–179, 2013.

[65] G. Zhang and Y. Wang, "Robust 3d face recognition based on resolution invariant features," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 1009–1019, 2011.

[66] E. Vezzetti, F. Marcolin, and G. Fracastoro, "3d face recognition: An automatic strategy based on geometrical descriptors and landmarks," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1768–1776, 2014.

[67] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, and N. Dagnes, "3d geometry-based automatic landmark localization in presence of facial

occlusions," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 14 177–14 205, 2018.

[68] C. Samir, A. Srivastava, and M. Daoudi, "Three-dimensional face recognition using shapes of facial curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1858–1863, 2006.

[69] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 80–95, 2009.

[70] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3d face recognition using elastic radial curves," 2010.

[71] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.

[72] X. Li and F. Da, "Efficient 3d face recognition handling facial expression and hair occlusion," *Image and Vision Computing*, vol. 30, no. 9, pp. 668–679, 2012.

[73] L. Ballihi, B. B. Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3-d-geometric features for efficient face recognition and gender classification," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1766–1779, 2012.

[74] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.

[75] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3d face recognition approach using local geometrical signatures," *Pattern Recognition*, vol. 47, no. 2, pp. 509–524, 2014.

[76] S. Berretti, A. Del Bimbo, and P. Pala, "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, pp. 374–389, 2012.

[77] F. R. Al-Osaimi, "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, pp. 658–672, 2015.

[78] M. Emambakhsh and A. Evans, "Nasal patches and curves for expression-robust 3d face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 995–1007, 2016.

[79] A. Abbad, K. Abbad, and H. Tairi, "3d face recognition: Multi-scale strategy based on geometric and local descriptors," *Computers & Electrical Engineering*, vol. 70, pp. 525–537, 2018.

[80] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: A quantum mechanical approach to shape analysis," in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2011, pp. 1626–1633.

[81] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[82] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3d face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190–2198, 2013.

[83] L. Shi, X. Wang, and Y. Shen, "Research on 3d face recognition method based on lbp and svm," *Optik*, vol. 220, p. 165157, 2020.

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

[85] N. Werghi, C. Tortorici, S. Berretti, and A. Del Bimbo, "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, pp. 964–979, 2016.

[86] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3d face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition*, vol. 46, no. 1, pp. 24–37, 2013.

[87] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin, "Covariance descriptors for 3d shape matching and retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4185–4192.

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

[89] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, "3-d face recognition using curvelet local features," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 172–175, 2013.

[90] Y. Ming, "Robust regional bounding spherical descriptor for 3d face recognition and emotion analysis," *Image and Vision Computing*, vol. 35, pp. 14–22, 2015.

[91] S. Soltanpour and Q. J. Wu, "High-order local normal derivative pattern (lndp) for 3d face recognition," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 2811–2815.

[92] S. Soltanpour and Q. M. J. Wu, "Weighted extreme sparse classifier and local derivative pattern for 3d face recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3020–3033, 2019.

[93] Y. Yu, F. Da, and Y. Guo, "Sparse icp with resampling and denoising for 3d face verification," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1917–1927, 2018.

[94] L. Spreeuwers, "Fast and accurate 3d face recognition," *International journal of computer vision*, vol. 93, no. 3, pp. 389–414, 2011.

[95] O. Ocegueda, G. Passalis, T. Theoharis, S. K. Shah, and I. A. Kakadiaris, "Ur3d-c: Linear dimensionality reduction for efficient 3d face recognition," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–6.

[96] Y. Ming and Q. Ruan, "Robust sparse bounding sphere for 3d face recognition," *Image and Vision Computing*, vol. 30, no. 8, pp. 524–534, 2012.

[97] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-d face recognition," *IEEE transactions on image processing*, vol. 22, no. 3, pp. 914–925, 2012.

[98] Y. Taghizadegan, H. Ghassemian, M. Naser-Moghaddasi *et al.*, "3d face recognition method using 2dpca-euclidean distance classification," *ACEEE International Journal on Control System and Instrumentation*, vol. 3, no. 1, pp. 1–5, 2012.

[99] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3d face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 2, pp. 381–397, 2012.

[100] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3d face recognition," *Neurocomputing*, vol. 129, pp. 445–457, 2014.

[101] N. I. Ratyal, I. A. Taj, U. I. Bajwa, and M. Sajid, "3d face recognition based on pose and expression invariant alignment," *Computers & Electrical Engineering*, vol. 46, pp. 241–255, 2015.

[102] Y. Tang, X. Sun, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "3d face recognition with asymptotic cones based principal curvatures," in *2015 International Conference on Biometrics (ICB)*. IEEE, 2015, pp. 466–472.

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

[104] M. Peter, J.-L. Minoi, and I. H. M. Hipiny, "3d face recognition using kernel-based pca approach," in *Computational Science and Technology*. Springer, 2019, pp. 77–86.

[105] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "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, pp. 1938–1951, 2011.

[106] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "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, pp. 1551–1565, 2012.

[107] N. Alyüz, B. Gökberk, L. Spreeuwers, R. Veldhuis, and L. Akarun, "Robust 3d face recognition in the presence of realistic occlusions," in *2012 5th IAPR International Conference on Biometrics (ICB)*. IEEE, 2012, pp. 111–118.

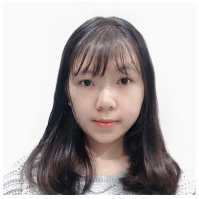
[108] N. Alyuz, B. Gokberk, and L. Akarun, "3-d face recognition under occlusion using masked projection," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 5, pp. 789–802, 2013.

[109] H. Fadaifard, G. Wolberg, and R. Haralick, "Multiscale 3d feature extraction and matching with an application to 3d face recognition," *Graphical Models*, vol. 75, no. 4, pp. 157–176, 2013.

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

[111] Bagchi, Parama, D. Bhattacharjee, and M. Nasipuri, "3d face recognition using surface normals," in *TENCON 2015-2015 IEEE Region 10 Conference*. IEEE, 2015, pp. 1–6.

[112] Y. Liang, Y. Zhang, and X.-X. Zeng, "Pose-invariant 3d face recognition using half face," *Signal Processing: Image Communication*, vol. 57, pp. 84–90, 2017.
[113] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[114] D. Kim, M. Hernandez, J. Choi, and G. Medioni, "Deep 3d face identification," in *2017 IEEE international joint conference on biometrics (IJCB)*. IEEE, 2017, pp. 133–142.
[115] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.
[116] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, 2014, pp. 1988–1996.
[117] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
[118] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
[119] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2746–2754.
[120] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1891–1898.
[121] Sun, Yi, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2892–2900.
[122] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.
[123] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
[124] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. Ieee, 2009, pp. 296–301.
[125] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: A 3d facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2013.
[126] J. D'Errico, "Surface fitting using gridfit," *MATLAB central file exchange*, vol. 643, 2005.
[127] S. Lin, F. Liu, Y. Liu, and L. Shen, "Local feature tensor based deep learning for 3d face recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.
[128] Y. Tan, H. Lin, Z. Xiao, S. Ding, and H. Chao, "Face recognition from sequential sparse 3d data via deep registration," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8.
[129] J. Feng, Q. Guo, Y. Guan, M. Wu, X. Zhang, and C. Ti, "3d face recognition method based on deep convolutional neural network," in *Smart Innovations in Communication and Computational Sciences*. Springer, 2019, pp. 123–130.
[130] E. C. Olivetti, J. Ferretti, G. Cirrincione, F. Nonis, S. Tornincasa, and F. Marcolin, "Deep cnn for 3d face recognition," in *International Conference on Design, Simulation, Manufacturing: The Innovation Exchange*. Springer, 2019, pp. 665–674.
[131] G. Ciravegna, G. Cirrincione, F. Marcolin, P. Barbiero, N. Dagnes, and E. Piccolo, "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*. Springer, 2020, pp. 223–233.
[132] K. Xu, X. Wang, Z. Hu, and Z. Zhang, "3d face recognition based on twin neural network combining deep map and texture," in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*. IEEE, 2019, pp. 1665–1668.
[133] Z. Zhang, F. Da, and Y. Yu, "Data-free point cloud network for 3d face recognition," *arXiv*, pp. arXiv–1911, 2019.
[134] M. Lüthi, T. Gerig, C. Jud, and T. Vetter, "Gaussian process morphable models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 8, pp. 1860–1873, 2017.
[135] G. Mu, D. Huang, G. Hu, J. Sun, and Y. Wang, "Led3d: A lightweight and efficient deep approach to recognizing low-quality 3d faces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5773–5782.
[136] A. R. Bhople, A. M. Shrivastava, and S. Prakash, "Point cloud based deep convolutional neural network for 3d face recognition."
[137] C. Cao, M. R. Swash, and H. Meng, "Reliable holoscopic 3d face recognition," in *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2020, pp. 696–701.
[138] K. Dutta, D. Bhattacharjee, and M. Nasipuri, "Sppcanet: a simple deep learning-based feature extraction approach for 3d face recognition," *Multimedia Tools and Applications*, pp. 1–24, 2020.
[139] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.



Yaping Jing received the Bachelor degree (Hons) of Information Technology from Deakin University, Australia in 2016. She is currently a PhD candidate in 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) at Deakin University, Australia. He spent more than two years working as a Research Fellow in Singapore. Prior to that, he received his Ph.D from Zhejiang University (China) in June 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. More information can be found at `http://www.xuequanlu.com`.

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