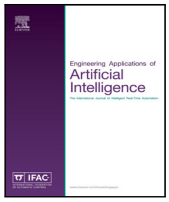
Engineering Applications of Artificial Intelligence 110 (2022) 104669

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Survey Paper

A comprehensive survey on 3D face recognition methods

Menghan Li *, Bin Huang, Guohui Tian

School of Control Science and Engineering, Shandong University, Shandong, China

ARTICLE INFO

Keywords:
3D face recognition
Deep learning
Expression
Pose
Occlusion
Survey

ABSTRACT

3D face recognition (3DFR) has emerged as an effective means of characterizing facial identity over the past several decades. Depending on the types of techniques used in recognition, these methods are categorized into traditional and modern. The former generally extract distinctive facial features (e.g. global, local, and hybrid features) for matching, whereas the latter rely primarily on deep learning to perform 3DFR in an end-to-end way. Many literature surveys have been carried out reviewing either traditional or modern methods alone, while only a few studies are conducted simultaneously on both of them. This survey presents a state-of-the-art for 3DFR covering both traditional and modern methods, focusing on the techniques used in face processing, feature extraction, and classification. In addition, we review some specific face recognition challenges, including pose, illumination, expression variations, self-occlusion, and spoofing attack. The commonly used 3D face datasets have been summarized as well.

1. Introduction

Face recognition (FR) is a technology that uses the face as the principle of biological characteristic for identity recognition. Compared with widely used biometrics, e.g. fingerprints, iris, and DNA, the face has great application perspectives with its reliability, security, and high social acceptance. FR has been extensively studied in the 2D domain, while 2DFR is sensitive to pose, illumination, expression (PIE) and self-occlusion. 3DFR can make full use of reliable facial geometry information to overcome the above limitations, hence, it has made great strides over the past few years. The main surveyed 3DFR methods are presented in Fig. 1.

During the last decades, there are a number of excellent surveys in 3DFR. The majority of earlier surveys summarize the traditional methods and challenges in single modal either 3D data or 2D images. Sandbach et al. (2012) gave a comprehensive survey of static and dynamic 3D facial expression recognition. 3D mesh video facial expression recognition has also been discussed (Danelakis et al., 2015), which are categorized according to features and classifiers employed to analyze the dynamic 3D face. Patil et al. (2015) presented a survey of 3DFR techniques in terms of registration, feature extraction and matching, and also reviewed it with a special focus on expression variations and occlusion. Corneanu et al. (2016) conducted a general overview of RGB, 3D, thermal and multimodal approaches for facial expression analysis, with their history, trends, and affect-related applications being summarized as well. Soltanpour et al. (2017) only investigated 3D local feature-based methods for recognition. Zhou and Xiao (2018) reviewed the non-specific conditions into three different categories: pose, expression, and occlusion, along with their limitations and solutions. Dagnes et al. (2018) reviewed the occlusion detection and restoration techniques for 3DFR. Alexandre et al. (2020) reviewed three main aspects of 3D facial expression recognition: face representation, preprocessing, and feature classification by investigating 49 specialized studies.

However, traditional methods may not suitable for unconstrained conditions, some of the recent review papers place a higher value on deep learning used for 3DFR. A survey by Guo and Zhang (2019) reviewed both the deep architectures and specific recognition problems, apart from the common variant PIE, video, 3D, and heterogeneous FR are addressed as well. Li and Deng (2020) presented a detailed review of static and dynamic 3D face expression recognition based on deep learning. Deep FR has also been surveyed by Wang and Deng (2021), different from (Guo and Zhang, 2019), this survey categorized face processing into two classes: "one-to-many augmentation" and "many-to-one normalization", with deep network architecture and loss function being summarized as well. Guha (2021) reported the journey from traditional to modern deep learning methods for both 2D and 3D FR.

The above surveys only review either traditional or modern methods, and the introduction and challenges in 3DFR, with less emphasis on face processing and recognition techniques. Research under unconstrained conditions is not fully discussed as well. Given the gap of the previous studies, in this paper, we present a comprehensive survey of 3DFR studies covering both traditional and modern methods along with their limitations and benefits, and mainly focus on the recognition

* Corresponding author.
E-mail addresses: menghan1220@mail.sdu.edu.cn (M. Li), huangbin@sdu.edu.cn (B. Huang), g.h.tian@sdu.edu.cn (G. Tian).

https://doi.org/10.1016/j.engappai.2022.104669
Received 4 August 2021; Received in revised form 31 October 2021; Accepted 10 January 2022
Available online 1 February 2022
0952-1976/

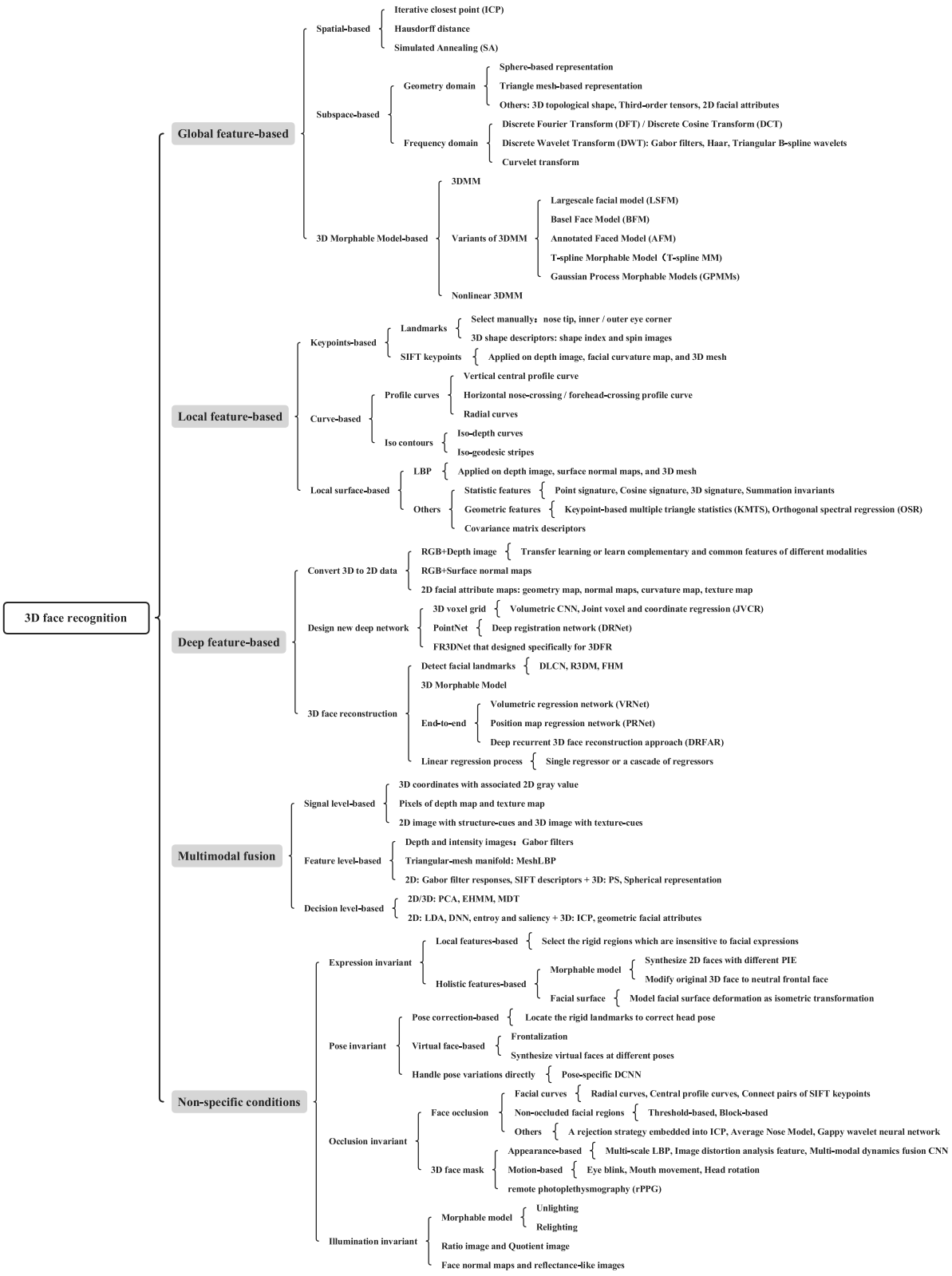


Fig. 1. The main surveyed 3DFR methods.

techniques used in face processing, feature extraction and classification. We also review some FR challenges (e.g. variant PIE, self-occlusion, and spoofing attack) and discuss research directions in the future. The main existing 3D face datasets are summarized as well.

The remainder of this survey is organized as follows: In Section 2, we introduce each module of 3DFR and summarize commonly used 3D face datasets. Sections 3, 4, 5 provides a comprehensive survey of global features, local features, and deep learning-based methods respectively. Section 6 surveys multimodal methods based on the fusion of 3D faces and 2D images. Section 7 surveys researches for 3DFR under unconstrained conditions, including variant PIE, self-occlusion, and spoofing attack. Section 8 discusses our surveyed methods and

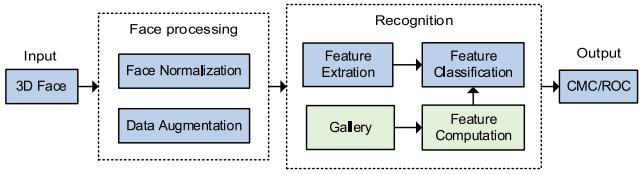


Fig. 2. A basic framework for 3DFR system.

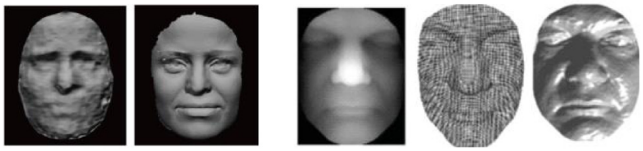


Fig. 3. (a) Face acquisition: 3D faces captured by Kinect (left) and Minolta scanner (right) (b) Face representation: depth image, 3D point cloud, and 3D mesh (from left to right).

presents the main challenges for 3DFR in the future. Finally, Section 9 gives the conclusion.

2. Related works

A basic framework for 3DFR system is shown in Fig. 2. Given the independence of 3D data acquisition, the existing 3DFR system typically involves face processing, feature extraction and classification. Before feeding 3D face into FR system, face anti-spoofing is significant to judge whether the input is the real face or people with 3D face mask. Face processing techniques are applied to the real input face as needed, then high discriminative features, either global or local features, are extracted from the preprocessed face for classification.

2.1. Face acquisition and representation

3D data acquisition is the first step in 3DFR, where the quality of acquired data and required time play a significant role in FR. There are two main types of acquisition techniques to capture 3D faces: active and passive ranging. In the active ranging techniques, like structured light methods, which actively projects one or more encoded light patterns onto the scene and measure the deformation on the objects' surfaces, e.g. Minolta Vivid 900/910 series, Inspeck Mega Capturor II. In the passive ranging techniques, like photometric stereo and multi-view stereo. The former passively accepts a set of images of the object under different illuminations to estimate the orientation field of 3D facial surface, while the latter employs multiple cameras placed at various known viewpoints from the subject to find corresponding points, e.g. 3DMD digitizer and Di3D.

Although 3D face scans obtained by the above acquisition techniques are of high quality, these types of data acquisition are costly and time-consuming. Recently, with the popularity of low-quality RGB-D cameras, e.g. Kinect and Realsense, it is capable of capturing the 3D face at a fast speed which makes its generalization on a larger scale feasible. However, the accuracy of obtained RGB-D image is comparatively low (see Fig. 3(a)), which requires additional preprocessing before fed to recognition system. Fig. 3(b) shows three main types of captured 3D data: 3D point cloud, 3D mesh, and depth image, all of which can be transformed into each other.

2.2. Face processing

Among the surveyed studies, face processing is an essential part before recognition, which can be categorized as "face normalization" and "data augmentation". The former aligns the face images of arbitrary

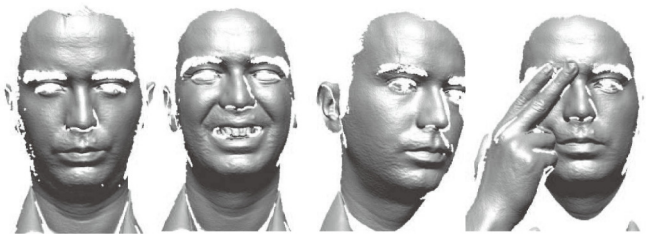


Fig. 4. Some degradation conditions in 3DFR: normal, expressions, head pose, and self-occlusion (from left to right).

view to extract discriminant facial features, while the latter generates a large number of virtual images from a single 2D image to overcome the insufficient of training data.

Given a set of face images, it is vital to detect face and remove nonface areas, then one or many preprocessing techniques, e.g. pose correction, face alignment, smoothing and hole filling, are applied to the detected face for the meaningful feature extraction. Some face alignment detectors that widely used in FR are Active Appearance Model (AAM), cascaded regression, and deep learning, in which deep learning are the most popular (Bellil et al., 2016; Chang et al., 2018; Le and Kakadiaris, 2019; Zhu et al., 2015). Besides, variations in expression, head pose, illumination and occlusion still impair the recognition performance (see Fig. 4), special normalization methods are required to align 3D face under these degradation conditions (see Section 7).

Deep learning network needs massive training data for feature learning, while most currently available datasets are insufficient to ensure generalizability of recognition task. Hence, data augmentation emerges as times required, which synthesizes many virtual faces at desired PIE via 3D rendering or other deep learning methods (Chung et al., 2020; Deng et al., 2018; Tran et al., 2017; Xu et al., 2021; Zhang et al., 2018a). Online data argumentation usually embeds random crop and horizontal flip in deep learning toolkits to enlarge the training dataset, while for offline data argumentation, apart from the most widely used 3DMM, some deep learning methods, e.g. CNN, Autoencoder, and GAN, have also been introduced to synthesize 2D faces to expand data on both size and diversity.

2.3. Feature extraction and classification

The extracted features are categorized as global feature, local feature, and deep feature. Global features mainly include geometry features, frequency features, 3DMM and its variants, while for local features, the keypoints, curves, and local surface of the face are usually extracted to form facial feature vector. Deep learning has recently of great interest in 3DFR, whose network architecture and loss function greatly affect the recognition performance. The typical Convolutional Neural Network (CNN) architecture (e.g. VGGNet, AlexNet, GoogleNet, and ResNet) are widely used as baseline models, while Autoencoder, Generative Adversarial Network (GAN), and Recurrent Neural Network (RNN) are also introduced for FR. Besides, many researchers focus on creating novel loss function to make features separable and discriminative, e.g. contrastive loss, triplet loss, center loss, and variants of Softmax loss.

As the final step, feature classification is a critical aspect for the evaluation of FR. Most of traditional methods may choose Support Vector Machines (SVMs), while Linear Discriminant Analysis (LDA), Hidden Markov Models (HMMs), Nearest-neighbor (NN), Random Forests are less frequently adopted in the surveyed studies. Feature extraction and classification are independent in traditional methods, while deep learning methods usually embed the feature classification in neutral networks to perform 3DFR in an end-to-end way.

Table 1
3D face datasets, P = pose, E = expression, I = illumination, O = occlusion, R = rotation, V = video.

Reference/Name	Data type	Intensity image	Subjects	Images	Elicit	URL
(2005)/FRGC v2.0	Depth image	Yes	466	4007	P, E, O	https://cvrl.nd.edu/projects/data/#face-recognition-grand-challenge-frgc-v20-data-collection
(2005c)/ND-Collection D	Depth image	Yes	277	953	E	https://cvrl.nd.edu/projects/data/#nd-collection-d
(2006)/MSU	Depth image	No	90	533	E	https://libguides.lib.msu.edu/eresources
(2008)/SHREC08	Depth image	No	61	427	P, E	https://engineering.purdue.edu/PRECISE/shrec08
(2011)/SHREC11	Depth image	No	130	780	P, O	https://www.nist.gov/itl/iad/shrec-2011-shape-retrieval-contest-range-scans
(2010)/Texas 3DFR D	Depth image	Yes	118	1149	E	http://live.ece.utexas.edu/research/texas3DFR/
(2011)/UMB-DB	Depth image	Yes	143	1473	E, O	http://www.ivl.disco.unimib.it/minisites/umbdb/
(2007)/ND-2006	Depth image	Yes	888	13450	E	https://cvrl.nd.edu/projects/data/#nd-2006-data-set
(2011)/3D-TEC	Depth image	Yes	214	428	E	https://cvrl.nd.edu/projects/data/#3d-twins-expression-challenge-3d-tec-data-set
(2008)/BosphorusDB	3D point cloud	Yes	105	4652	P, E, O	http://bosphorus.ee.boun.edu.tr/
(2006)/BU-3DFE	Mesh	Yes	100	2500	E	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
(2008)/BU-4DFE	3D video	Yes	101	60600	E	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
(2008)/UOY	Mesh	Yes	350	5000	P, E	http://www.cs.york.ac.uk/~tomh/
(2006)/ZJU_3DFED	Mesh	Yes	40	360	E	http://www.bjut
(2009)/BJUT-3D	Mesh	Yes	500	500	–	http://www.bjpu.edu.cn/sci/multimedia/mul-lab/3dface//facedatabase.htm
(2006)/FRAV3D	Mesh	Yes	105	–	P	http://www.frav.es/databases/FRAV3d/
(2020)/FaceScape	Mesh	Yes	938	18760	E	https://facescape.nju.edu.cn/
(2011)/ICT-3DREF	Mesh	Yes	23	–	E, I	https://vgl.ict.usc.edu/Data/3DRFE/
(2014)/UHDB11	Mesh	Yes	23	3312	P, I	http://cbl.uh.edu/URxD/datasets
(2017)/UHDB31	Mesh	Yes	77	25872	P, I, R	http://cbl.uh.edu/repository-data
(2012)/VAP RGB-D	Depth image	Yes	31	151	P, E	https://vap.aau.dk/rgb-d-face-database/
(2014)/FaceWareHouse	Depth image	Yes	150	3000	E	http://kunzhou.net/zjugaps/facewarehouse/
(2016)/Lock3DFace	Depth image	Yes	509	5711	P, E	http://irip.buaa.edu.cn/lock3dface/index.html
(2014)/IIIT-D	Depth image	Yes	106	4605	P, E	http://www.iab-rubric.org/resources.html
(2014)/KinectFaceDB	Depth image	Yes	52	–	P, E, O, V	http://rgb-d.eurecom.fr/

Performance measures usually considered for matching scores are Receiver Operating Characteristic (ROC) and Cumulative Match Characteristic (CMC), applied to face verification and face recognition respectively. The former needs to compare the test face with each face in the gallery to find the best match, while the latter only needs to confirm or reject the claimed identity of the test face. The metrics used in the surveyed studies consist of Rank-1 Recognition Rate (RR1), Verification Rate (VR) and Equal Error Recognition (EER).

2.4. 3D face datasets

Over the past few years, there are numerous 3D face datasets have been set up for FR. This section summarizes the main existing 3D face datasets, including both high-accuracy datasets, acquired from 3D scanners and low-accuracy datasets, acquired from RGB-D cameras. Table 1 describes our surveyed datasets under four different categories, e.g. data type, intensity image, number of subjects and images, and indicates some non-specific conditions contained for each.

3. Global feature-based methods

Focusing on the global similarity of faces, global feature-based methods usually transform the whole 3D face into different representations, from which extract a set of global features for matching. This section surveys and explains the main existing global feature-based methods and groups them into three different categories: spatial-based, subspace-based, and morphable model-based methods.

3.1. Spatial-based methods

Spatial-based methods usually take coordinates or normal vectors of 3D point cloud for face matching without extracting facial features.

3.1.1. Iterative closest point (ICP)

As the baseline algorithm for 3D face matching, the ICP algorithm is first introduced by Besl and McKay (1992), which iteratively aligns two 3D faces and uses the matching error between two cloud points for recognition. However, ICP is not suitable for surfaces with non-rigid deformation, such as facial expression variations. An effective modification is that only used ICP to the rigid regions of face, in other words, the weight of undeformed regions should be increased while that of other regions reduced accordingly. Chang et al. (2005a) used surface curvature to segment the face and then extract the nose region for ICP registration. An extension of this work is proposed in (2006), which selected multiple overlapping regions around the nose and fused matching scores as the final result. A partial ICP method is proposed by Wang et al. (2006), which modified the selection rules of initial points and calculation rules of associated points to extract facial rigid parts under expression variations.

In addition, there are other modified ICP algorithms. Papatheodorou and Rueckert (2004) proposed a novel 4D data composed of 2D gray value and 3D coordinators to represent 3D face. The ICP algorithm is then applied for registration. The Iterative Closest Normal Point (ICNP) method is proposed by Mohammadzade and Hatzinakos (2013) to register the nose region, which denotes the sampled points as the closest normal points for recognition. Experiments show that the surface normal vectors contain more discriminative information than point coordinates. A rejection strategy is embedded into ICP to eliminate the impacts of occlusion (Liu et al., 2012). Besides, the facial surface is described as the Spherical Depth Map (SDM) to remove the redundant vertices, thus reducing the time consumed by ICP operation. Lin et al. (2014) integrated the elastic deformation into non-rigid ICP, where the radial basis function is used to deform the template face to make it more similar to the target face. Cheng et al. (2015) used the active non-rigid ICP algorithm to register the part-based 3D face statistical model. Yu et al. (2019) proposed a sparse ICP with resampling and denoising, which integrated sparsity-inducing norms to reduce the sensitivity to noise, outliers, and missing data.

Table 2
Spatial-based methods.

Reference	Main algorithm
Besl and McKay (1992)	First introduce ICP
Chang et al. (2005a)	Extract nose region for ICP
Chang et al. (2006)	Select multiple overlapping regions around nose for ICP
Wang et al. (2006)	Partial ICP
Papatheodorou and Rueckert (2004)	3D coordinators and associated 2D gray value for ICP
Mohammadzade and Hatzinakos (2013)	Iterative Closest Normal Point (ICNP)
Liu et al. (2012)	A rejection strategy embedded into ICP
Lin et al. (2014)	Integrate elastic deformation into non-rigid ICP
Yu et al. (2019)	Sparse ICP with resampling and denoising
Li et al. (2010)	Least trimmed square Hausdorff distance
Queirolo et al. (2010)	Simulated Annealing (SA)

3.1.2. Hausdorff distance

After performing ICP, the average distance between two cloud points is calculated as the similarity measure. Since its vulnerability to noise, some studies chose Hausdorff distance and its variants. Li et al. (2010) used the central profile alignment based on the least trimmed square Hausdorff distance to form a rejection classifier for eliminating the candidate faces. Then, the modified ICP algorithm is applied to register rigid regions of the remaining faces. Apart from the ICP and Hausdorff distance, Queirolo et al. (2010) proposed Simulated Annealing (SA) algorithm for depth image registration, which combines coarse alignment (MLESAC) with fine registration (SIM) to obtain the authentication score of each facial part, e.g. the entire face, forehead, circular and elliptical areas around the nose.

Table 2 summarizes our surveyed spatial-based methods. The main limitation of these methods is their lack of robustness for 3DFR under unconstrained conditions, so that they can only recognize the frontal face without expression. As a global iterative algorithm, they are not suitable for large-scale 3D face datasets since the high computational cost and time-consuming. All of these disadvantages limit them only used for face matching rather than FR independently.

3.2. Subspace-based methods

Subspace-based methods can be described as mapping 3D face into different attribute maps for feature extraction and classification, which are classified into geometry domain-based and frequency domain-based methods according to different feature extraction domains.

3.2.1. Geometry domain-based

Sphere representation. Pan et al. (2005) located the nose tip by detecting the facial bilateral symmetry plane, then extracted the intersection region of facial surface and spherical surface centered at nose tip as the region of interests. The ROIs are further triangulated and parameterized into an isomorphic 2D planar circle with the associated depth values being mapped as well. Mian et al. (2007) converted the 3D point cloud to a spherical representation centered at the nose tip, then employed SIFT descriptor to form a rejection classifier for eliminating candidates' faces, and ICP to match eyes-forehead region and nose region of remaining faces. Liu et al. (2013) first represented 3D point cloud as spherical depth map, where the center of the fitting sphere rather than the center of mass is at the origin, by which spherical harmonic features can be calculated.

Ming and Ruan (2012) projected the preprocessed 3D point cloud on the bounding spheres with the center being the face centroid, where the ratio between the distances of points and its corresponding centroid and radius of the bounding sphere is referred to as the value of the spherical point, called bounding sphere representation (BSR). A robust group sparse regression model is then proposed to extract the intrinsic discriminant features. An extension of this framework is proposed in (2015), with facial shape index and spherical bands being employed, each 3D point cloud is divided into a group of facial regions, on which regional bounding sphere descriptors (RBSR) are calculated respectively. A region and global regression mapping technology is employed to the weighted regional descriptor for accurate classification.

3D mesh. Apart from the sphere representation of 3D face, Xu et al. (2004) fitted the 3D point cloud to a regular mesh by using hierarchical mesh method, then global geometric features described as the depth value of each mesh point, and local shape variation information extracted on some rigid regions of the face, are connected as a feature vector to characterize the individual. Triangle mesh-based representation has also been employed by Li and Zhang (2007). The multiple geometric attributes, including angles, geodesic distances, mean and Gaussian curvatures, are calculated on 3D mesh, where the component-wise weights depend on the degree of individual attributes affected by expression variations. Li et al. (2009) proposed the sparse representation based on a uniform remeshing scheme, where a feature pooling and ranking scheme is designed to improve the robustness to expression variations.

Others. There are many other approaches proposed to represent 3D face. The canonical face depth map (CFDM), a standardized 3D face representation, is proposed by Colbry and Stockman (2007). After detecting the mid-line plane for face alignment, the parabolic cylinder is fitted to the face to establish an orthogonal face-based coordinate, from which a normalized feature vector is extracted for recognition. 3D topological shape information is also used together with geometric information to represent the 3D face (Samir et al., 2005), where the mean curvature of topologically connected components is calculated as the facial recognition attributes.

Mian et al. (2005) represented the 3D face as a set of third-order tensors indexed by a 4D hash table. During recognition, the voting-based method is applied to rank the tensors in the gallery, while the similarity is calculated only for the high-ranked faces. The face with the highest similarity is selected as the recognized face. The tensor feature has also been employed by Al-Osaimi et al. (2008). Multiple global and local rank-0 tensors are calculated to form rank-0 tensor fields that are invariant to rigid transformations, then the local and global tensor fields are combined to form a 2D histogram for recognition.

Zhang et al. (2012) represented each 3D facial surface as a 2D attribute image. Specifically, each 3D facial surface is mapped homomorphically onto a 2D lattice by using discrete conformal mapping, where the local mean curvature of each vertex is used to represent the value of associated site. The 2D lattice is then interpolated to generate a 2D attribute image. Based on this model, the sparse representation is applied to these 2D attribute images for recognition (Guo et al., 2013)

3.2.2. Frequency domain-based

Fourier analysis. Discrete Cosine Transform (DCT), regarded as the Discrete Fourier Transform (DFT) in the real number field, is often used to compress the lossy data in the process of signal and image. Ekenel et al. (2007) employed the closest-point mapping and the ray-casting mapping to construct depth images, all of which are divided into multiple non-overlapping regions where DCT performs to extract local information. Besides hardly handling the abrupt and non-stationary signals, the Fourier analysis is incapable of describing the local information of signals in the time domain.

Wavelet analysis. As a local transform of space and frequency, wavelet analysis emerges as the times required. The Barycentric wavelet kernels (BWKs) is employed to decompose a hexagonal region centered on the nose tip, with triangular B-spline wavelets being selected as the adopted wavelet (Jahanbin et al., 2007). The PCA is applied to reduce the dimension of extracted coefficients at each resolution. Cook et al. (2007a) used Log-Gabor filters, DCT, and Discrete Wavelet Transform (DWT) to extract facial features, with Haar wavelet being selected as the adopted wavelet. Experiments demonstrate that Log-Gabor filters have a better recognition performance than others. Xu et al. (2009)

Table 3
Subspace-based methods.

Reference	Main algorithm
Mian et al. (2007)	Spherical Face Representation (SFR)
Liu et al. (2013)	Spherical depth map (SDM)
Ming and Ruan (2012)	Bounding sphere representation (BSR)
Ming (2015)	Regional bounding sphere descriptors (RBSR)
Li and Zhang (2007)	Triangle mesh-based representation
Li et al. (2009)	The sparse representation based on uniform remeshing
Samir et al. (2005)	3D topological shape
Mian et al. (2005)	The 3th tensors indexed with 4D hash table
Al-Osaimi et al. (2008)	Global and local rank-0 tensors
Zhang et al. (2012)	2D facial lattice
Ekenel et al. (2007)	Discrete Cosine Transform (DCT)
Jahanbin et al. (2007)	Barycentric wavelet kernels (BWKs)
Cook et al. (2007a)	Log-Gabor filters, DCT, DWT
Xu et al. (2009)	Local Gabor features
Wang et al. (2010)	Haar-like, Gabor, LBP
Elaiwat et al. (2013, 2014)	Curvelet transform

extracted local Gabor features from depth and intensity images respectively, with AdaBoost being used to select the most efficient features for classification. A collective shape difference classifier (CSDC) is proposed by Wang et al. (2010). Specifically, a signed shape difference map is computed between two aligned 3D faces, from which Haar-like feature, Gabor feature, and LBP are extracted at different positions. The most discriminative local features are chosen by a boosting algorithm and then trained as weak classifiers for assembling three collective strong classifiers. Finally, the most efficient CSDC, Haar feature-based, is selected for recognition.

Curvelet transform. In reality, wavelet analysis is no longer suitable for feature extraction under the high dimensional situation, so how to make full use of the high-dimensional geometric characteristics of 3D face attracts the attention of many scholars. Elaiwat et al. (2014) first introduced Curvelet transform to detect facial salient points and calculated the curvelet coefficients and orientations of different scales, then the local surface descriptor is constructed around each detected keypoint to extract highly descriptive rotation-invariant features. Curvelet transform has also been used in (2013), where the curvelet features are employed to describe semi-rigid regions of 2D and 3D faces, e.g. nose and eyes-forehead regions. The final recognition results are obtained by fusing multiple matching scores at the decision level.

Table 3 summarizes the subspace-based methods presented in this section. Geometry domain-based methods usually map 3D point cloud into sphere representation or regular mesh to yield robust descriptors, while the latter usually transforms the depth image into frequency domain via Fourier transform, wavelet transform or curvelet transform, from which the related frequency features are extracted for the following recognition.

3.3. 3D morphable model-based methods

As the morphable model can effectively encode the geometric and texture variations of the face, 3DMM and its variants have been widely used to reconstruct the 3D face, which typically involves two major steps: model construction and face fitting. This section summarizes the 3DMM and its variants, including conventional and nonlinear 3DMM, as well as some new morphable models.

3.3.1. 3D morphable model (3DMM)

As a general face shape and texture representation, 3D morphable model (3DMM), which can be defined as the linear combinations of model bases, is commonly used to reconstruct 3D face model from one or more 2D images.

Blanz and Vetter (2003) initially introduced the 3DMM to estimate 3D facial shape and texture from a single 2D image, with the corresponding illumination, pose information, and camera parameters being judged at the same time. This 3DMM fitting approach has been followed by many others. Amberg et al. (2008) fitted the neutral and expressive 3DMM to 3D face scans separately, with non-rigid ICP being used for registration. Mpiperis et al. (2008) built an elastic deformation model to fit the target face and employed the bilinear models to decouple the influence of expression and identity. Seven morphable expression models and one morphable identity model are built to separate identity and expressions (ter Haar and Veltkamp, 2010). Al-Osaimi et al. (2009) built an expression deformation model by learning the shape residuals of expressive and non-expressive faces, while PCA is only used to model the facial expressions leaving out the interpersonal disparities. A largescale facial model (LSFM), constructed from 9663 distinct facial identities, is presented by Booth et al. (2018), which allows for the construction of not only global 3DMM but also specific models for age, gender, or ethnicity groups as well.

The RGB-D camera has the advantage of lower cost and higher speed on data acquisition, whereas the quality of the depth images obtained is comparatively lower. Therefore, it is essential to preprocess the depth images. Hsu et al. (2014) proposed a resurfacing approach to tackle the quantization noise on the depth image, further rendering the textured 3D face model to generate 2D face with specific poses. RGB-D images have been also employed to construct morphable expression model by Cao et al. (2014). Meyer and Do (2018) employed 3DMM to extract morphable parameters on low-quality depth images, where the point-to-point distance between the 3DMM and the depth image is selected as the similarity measure.

3.3.2. Variants of 3DMM

Although 3DMM has the advantage of being invariant to pose and expressions, it also has many limitations, such as long modeling time, high calculational cost, and modeling the same image with different results, which motivates researchers to study variants of 3DMM or new morphable models for 3D face reconstruction.

A statistical 3D shape and texture model, namely Basel Face Model (BFM), is proposed by Paysan et al. (2009), where the pose, lighting, and identity parameters are separated and only identity parameters will be selected for recognition. However, BFM can only model the neutral face since its sensitivity to facial expressions. Passalis et al. (2011) proposed Annotated Faced Model (AFM) to fit face scan. Each fitted model is converted to a geometry image and a normal image, on which Haar wavelet transform is applied for feature extraction. The similar work can be found in (Kakadiaris et al., 2007; Passalis et al., 2005; Theoharis et al., 2008).

Zou et al. (2007) selected a set of depth images as example faces and another depth image as a "generic face". The general face is wrapped to match each of the example faces in the least mean square sense, while the wrapped coefficients are extracted as facial features and fed into Mahalanobis distance-based classifier for recognition. The low-dimensional expression manifold is employed with the morphable neutral face model to reconstruct the 3D face model as well as expression deformation from a single face image (Wang and Lai, 2011). Gilani and Mian (2016) proposed a fully convolutional deep network to locate three facial landmarks (e.g. the left, right inner eyes, and nose tip) for face alignment. The seed points are then extracted by evolving level set curves with a single curvature dependent adaptive speed function. By matching the seed points between a template face and the target face, the morphable model with dense correspondence is generated, which is further registered and fitted to each probe face for the identity, expression, and pose parameters extraction.

Peng et al. (2017) proposed a parametric T-spline morphable model (T-spline MM) to represent 3D face. Specifically, C^2 T-spline is employed to describe the face surface, then T-mesh is divided into several shape units to refine the facial deformation. A fitting algorithm is proposed to approach the details of both identity and expression, even in the presence of occlusion. Luthi et al. (2018) proposed a generalization of point distribution models (PDMs), called Gaussian Process

morphable models (GPMMs). Any shape variation can be defined on the basis of Gaussian process, overcoming the limitation that PDMs can only model linear deformation. In addition, Gaussian process can be applied to combine different models to mimic more sophisticated registration schemes.

3.3.3. Nonlinear 3DMM

Conventional 3DMM is learned from 2D face images with associated 3D face scans, with PCA being employed to cope with facial shape and texture information. Limited by the type and amount of training data, the linear 3DMM may be incapable of reconstructing 3D face completely and accurately. Therefore, nonlinear 3DMM, learned from a large set of unconstrained 2D face images without collecting 3D face scans, is first introduced by Tran and Liu (2018). In particular, a network encoder is employed to estimate the relevant parameters of 2D face images, while two decoders, multi-layer perceptron (MLP) and CNN, are designed to extract shape and texture parameters respectively. The shape, texture, and camera projection parameters are fused to reconstruct the original input face.

Nonlinear 3DMM is optimized in (2019). On the one hand, strongly regularized proxies are used together with weakly regularized components to recover detailed shape and albedo. On the other hand, a dual-pathway network is proposed to decoder the shape and albedo, where the global pathway is built for inferring the global structure and the local pathway with four sub-networks is designed to model different facial parts (e.g. eyes, nose and mouth). The local pathways' inputs are fused as a single local feature that is simply concatenated with a global feature for the final recognition. In addition, the soft symmetry loss function is applied to occluded regions.

A set of in-the-wild face images, that is 2D face images captured under unconstrained conditions, are used to establish a nonlinear 3DMM in (2021). In-the-wild face images have also been applied by Tran et al. (2017), where the shape and texture parameters of the nonlinear 3DMM are directly regressed by pre-trained DCNN. Besides, a method for generating a large number of labeled examples based on Piotraschke and Blanz (2016) is proposed to overcome the insufficiency of training data. The classical nonlinear 3DMM uses DCNNs to regress the facial shape and texture parameters, which may result in a large network with many parameters. To address this problem, Zhou et al. (2019) proposed a joint texture and shape convolutional mesh auto-decoders, which can be used to train in-the-wild face images at the speed of 2500FPS.

Table 4 summarizes our surveyed 3DMM and its variants. These methods are robust to variant PIE, so they can be applied to unsupervised learning 3D models from large scale in-the-wild 2D images. Besides, most of data argumentation methods use 3DMM to generate a large number of virtual faces to handle the insufficiency of training data, hence, it is essential to distinguish the difference between synthetic images and real images.

Table 4
3DMM and its variants.

Reference	Main algorithm
Blanz and Vetter (2003)	3DMM
Amberg et al. (2008)	The neutral and expressive 3DMM
Mpiperis et al. (2008)	The elastic deformation model
ter Haar and Veltkamp (2010)	Seven morphable expression models and one morphable identity model
Al-Osaimi et al. (2009)	Expression deformation model
Booth et al. (2018)	LSFM
Meyer and Do (2018)	Employ RGB-D images to construct morphable expression model
Cao et al. (2014)	FaceWarehouse DB
Paysan et al. (2009)	Basel Face Model (BFM)
Passalis et al. (2011)	Annotated Faced Model (AFM)
Zou et al. (2007)	Warped Example Faces
Wang and Lai (2011)	Low-dimensional expression manifold and morphable neutral face model
Peng et al. (2017)	T-spline MM
Luthi et al. (2018)	Gaussian Process morphable models (GPMMs)
Tran et al. (2017)	Nonlinear 3DMM
Zhou et al. (2019)	Joint texture and shape convolutional mesh auto-decoders

4. Local feature-based methods

Local feature-based methods usually extract distinctive compact features as 3D local facial information, which can be categorized as keypoints-based, curve-based, and local surface-based methods.

4.1. Keypoints-based methods

3D keypoints are interest points of shape, which are detected according to some geometric information of the face surface. The keypoints-based methods usually include three steps: keypoints detection, feature description, and matching. This section reviews two kinds of keypoints, one is the landmarks based on facial organs, the other is the salient keypoints on the facial surface, such as Scale invariant feature transform (SIFT) keypoints.

4.1.1. Landmark-based methods

In these methods, the nose tip and eye corner are the key landmarks, while the landmarks from cheek, mouth, and other regions easily affected by expression variations are rarely utilized for recognition.

Lu et al. (2004) selected the inner eye, outer eye, and nose tip to coarsely align the 2.5D scan with the full 3D model, and then used hybrid ICP for fine registration. Gupta et al. (2007) manually located 25 facial fiducial points associated with anthropometric proportions on the color image, which performs better than selecting facial reference points arbitrarily. Emambakhsh and Evans (2017) detected seven keypoints on the nasal region and used Gabor-wavelet filtered depth maps to extract the nasal surface normal. A set of spherical patches and curves are located as nasal feature descriptors, in which the most stable are selected via a genetic algorithm to overcome expression variations. These methods usually extract geometry information in the neighborhood of the landmarks to generate reliable local features, rather than match the landmarks directly.

There are other proposed methods to detect landmarks, such as shape index. Perakis et al. (2013) used 3D shape descriptors including shape index and spin images to extract candidate landmarks, where the shape index is a continuous principal curvature map of the facial surface and spin images are local descriptors of facial 3D point distribution. The detected landmarks are matched with facial anatomical landmarks based on facial landmark model for identification. The shape index is also used with facial curves for landmark detection (Guo et al., 2017), there are two proposed keypoints detection methods, one is to sample uniformly on the valuable profile which is a nasal curve intercepted from the vertical central profile, the other is to screen the facial surface by using the mean curvature. The histograms of shape index, slant angles, and direction angles are concatenated as local features for matching. Guo and Da (2019) used the multiscale shape index to detect the facial landmarks. The 3D histogram of normal distributions on each keypoint is constructed to filter irrelevant candidate faces, while, curvature, shape index, Willmore energy, normal angle, distance to local plane, and the Euclidean distance from the reference to the nose tip, are calculated to generate covariance matrix descriptors to further match the remaining faces.

4.1.2. SIFT keypoints-based

Scale invariant feature transform (SIFT) is a keypoint detector. In this method, the extreme points in different scale spaces are detected as keypoints, around which the facial features and their directions are calculated as local feature descriptors. SIFT keypoints are initially introduced for FR (Mian et al., 2008). In this method, the facial surface

Table 5
Keypoints-based methods.

Reference	Main algorithm
Lu et al. (2004)	Inner/outer eye and nose tip
Gupta et al. (2007)	25 anthropometric facial fiducial points
Emambakhsh and Evans (2017)	Seven keypoints on the nose
Perakis et al. (2013)	3D shape descriptors
Guo and Da (2019)	Multiscale shape variation indexes
Guo et al. (2017)	Valuable profile and mean curvature
Mian et al. (2008)	3D tensor and 2D SIFT
Berretti et al. (2011)	SIFT applied to depth image
Inan and Halici (2012)	SIFT applied to facial curvature map
Smeets et al. (2013)	MeshSIFT
Li et al. (2011),	MeshSIFT
Li et al. (2015)	MeshSIFT
Berretti et al. (2014)	MeshDOG
Lin et al. (2019)	MeshSIFT

where shape variation is high is first identified as the keypoints, around which the tensor features are calculated as 3D features and SIFT features are employed in the 2D domain, all of them are fused at the feature and decision level respectively.

The SIFT-like keypoints can be detected from the depth image, facial curvature map, and 3D mesh. Berretti et al. (2011) extracted SIFT keypoints on the depth image and calculated facial features in spherical regions centered at each detected keypoint with different radius. The small radius can capture the details in small regions, while the large radius contains more discriminatory features. SIFT keypoints detection has also been applied in the facial curvature map, including shape index, curvedness, Gaussian and mean curvature values on each keypoint of 3D face (Inan and Halici, 2012).

As SIFT-like detector applied on the depth image or curvature image are sensitive to large pose variations, Smeets et al. (2013) directly extracted SIFT keypoints on 3D mesh, referred to as MeshSIFT, both histograms of shape index and slant angles are Gaussian weighted and concatenated as feature vectors. This MeshSIFT-like keypoints detector has also been used by Li et al. (2011), which used both maximum and minimum curvatures to detect local extrema within the 3D Gaussian scale space. Multiple order surface differential quantities, including surface gradient, shape index, and gradient of shape index, are calculated and concatenated to describe the local face surface. The extension work (Li et al., 2015) is presented, where a multi-task sparse representation based on fine-grained matching algorithm performs better than match two feature vectors through slant angle directly.

Considering the usage of generous SIFT keypoints results in a high computational cost, the MeshDOG keypoints detector and the multi-ring Geometric Histogram descriptor, an extension framework of (Berretti et al., 2013a), is proposed by Berretti et al. (2014). The detection of keypoints proceeds in three steps: scale-space, percentage threshold, and corner analysis, the normal and difference between the minimum and maximum perpendicular distance of two facets are calculated to create descriptors. Lin et al. (2019) also used this keypoints detection method on 3D mesh, and calculated 3rd order feature tensor of each salient keypoint to represent local features. Besides, the Voronoi diagram subdivision is proposed for data augmentation, which generates a large number of intra-personal and extra-personal similarity tensors to train the ResNet.

Table 5 summarizes surveyed keypoints-based methods. Besides creating feature vectors for FR, landmarks are widely used to correct pose for coarse alignment in the face processing, while the sparsity of keypoints may undermine the recognition performance. The SIFT-like keypoints are invariant to rotations, scaling, and illumination, however, numerous keypoints described by high dimensional feature vectors inevitably lead to higher computational cost, and the sensitivity to noisy data is worthy of consideration as well.

4.2. Curve-based methods

Curve-based methods use a set of representative curves over the surface to represent the face, making it not only less sparse than keypoints but also much easier than freeform surface matching. There are two types of commonly used curves: profile curves and iso contours.

4.2.1. Profile curves-based

Profile curves are defined as open curves that start and end on the edge of the face. Pan and Wu (2005) located the central profile curve by detecting the bilateral symmetry plane, with horizontal nose-crossing and forehead-crossing profile curves being employed for recognition. The curvature has also been used to extract the convex crest curves in convex local regions, with the central profile curve and horizontal nose tip-crossing profile curve being further located, these three types of curves are fused with appropriate weights at the feature level (Li et al., 2018).

Another specific kind of open curves are radial curves, from a center point to a face extremity. As the nose is less affected by expression variations and contains more distinctive shape information than other facial regions, the nose tip is generally selected as the center point of radial curves. Radial curves have been introduced by Drira et al. (2010) and extended in (2013). With the origin being the nose tip, a plane is used to segment the facial surface at a certain angle. Radial curves are defined as the intersection between the plane and the facial surface. The discontinuous or too short curves caused by occlusion will be discarded, while the remaining curves are applied to construct a Riemannian analysis framework for recognition.

The circular curves are also used together with radial curves to approximate the facial surface. The Adaboost algorithm is employed to select a minimal but effective set of curves for recognition, mostly existing in nose, forehead, and other relatively rigid regions. Radial curves have also been used by Lei et al. (2014), where the angular radial signatures (ARS) are defined as a set of curves, emanating from the nose tip at an interval. The kernel principal component analysis is used to project the ARS features along with different directions, where the depth value is computed as the value of ARS.

4.2.2. Iso contours-based

Iso contours are defined as level curves with different lengths and without intersections, which is generally defined at different heights below the nose tip. Iso-depth curves, firstly introduced by Samir et al. (2006), are obtained by translating a plane over the facial surface, that is a set of closed horizontal intersection curves between the facial surface and a plane. The geodesic distance defined in (2004) is used as a similarity measure between two corresponding curves. This framework is extended by Samir et al. (2009), the 3D face is represented as a set of level curves of a surface distance function, with the Riemannian analysis framework being employed to quantify the facial shape differences. The latter is invariant to facial rigid rotations and translations compared with the former (Samir et al., 2006). There are many other studies that used iso contours to represent the face. Lee and Krim (2017) extracted the deformed circular curves with simple constraints and calculated the shortest geodesic distances between a reference and points on the curve as the similarity measure, further reducing dimensionality by using the Fourier transform.

Motivated by the isometry assumption (Bronstein et al., 2005), a geodesic polar parameterization of the face surface is proposed by Mpiperis et al. (2007), which selected the nose tip as the geodesic pole, the radial curve from nose tip as the geodesic path, and the closed curve centered at nose tip as the geodesic circle. In addition, an active appearance model (AAM) is trained to track the boundary of the lips, which effectively handles the special case of an open mouth that violates the isometry assumption. Iso-geodesic stripes have also been applied by Berretti et al. (2010). The face is partitioned into eight iso-geodesic stripes, and arcs between pairs of iso-geodesic stripes,

Table 6
Curve-based methods.

Reference	Main algorithm
Pan and Wu (2005)	Central profile, horizontal nose-crossing and forehead-crossing profiles
Li et al. (2018).	Convex crest curves, central profile, and horizontal nose tip-crossing profile
Drira et al. (2010)	Radial curves
Ballihi et al. (2012)	Circular curves and radial curves
Lei et al. (2014)	Angular radial signatures (ARS)
Samir et al. (2006)	Iso-depth curves
Lee and Krim (2017)	Deformed circular curves with simple constraints
Mpiperis et al. (2007)	Geodesic polar representation
Berretti et al. (2010)	Iso-geodesic stripes
Gilani et al. (2017)	Geodesic level set curves

called 3D weighted walkthrough (3DWW), are calculated to obtain the relative spatial displacement. The graph-based algorithm, which has good invariance to facial expressions, is employed for face matching. Gilani et al. (2017) also used the iso-geodesic stripes to segment the 3D face into Voronoi regions and extracted the intrinsic features of different regions for matching elastically.

Table 6 summarizes our surveyed curve-based methods. The vertical central profile is the key curve, with radial curves and iso contours also being used for recognition. Most of curve-based methods select the nose tip as the reference point, however, pose changes, hair occlusion, and missing data may affect the correct detection of nose tip, further undermining the accuracy of face alignment and curve extraction.

4.3. Local surface-based methods

Local surface-based methods extract local geometric information from some patches or the area neighboring keypoints of the facial surface, which may be classified into LBP-based and other methods.

4.3.1. LBP-based

Local Binary Pattern (LBP), a local shape descriptor, which has the advantages of being invariant to rotation and gray, is initially introduced by Ojala et al. (2002) to extract the local texture features on 2D images. The common method is to extract LBP from local patches on depth image for FR, however, the keypoints on the depth image are too sparse to distinguish the face, the traditional LBP has been modified as times required. LBP descriptors can be calculated on depth image, 3D mesh, and surface normal map.

Global statistics of geometrical features are used together with local 3D LBP histograms features on depth images to approximate the facial surface (Huang et al., 2006). Huang et al. (2012) proposed a multiscale extended LBP, which encodes exact gray value differences between the central pixel and neighboring ones by embedding a multiscale scheme. In order to extract sufficient local features, SIFT is applied to detect the keypoints on LBP-based facial representations. Lv et al. (2015) proposed a region-based extended LBP to extract and encode facial information of different regions. Werghi et al. (2015b) used LBP to represent the texture of 3D mesh manifold. For each central facet on the mesh, three adjacent facets are considered edge facets of the central facet, then the facets around the edge facets are extracted as new edge facets and iterate the above procedure. Finally, a ring of ordered facets around the central facet can be generated. The mesh LBP has also been used in (2016), the vertical descriptor and horizontal descriptor are extracted on the mesh-LBP, and the two descriptors are further fused to obtain the vertical and horizontal descriptors, referred to as mesh-VHLBP.

More local shape information can be extracted using the surface normal instead of depth value. Tang et al. (2013) extracted and encoded the facial depth and normal information, then presented three statistical histogram matching strategies, the direct LBP, the weighted LBP, and the self-adaptive voting LBP for recognition. Li et al. (2014) proposed the multi-scale and multi-component local normal patterns (MSMC-LNP) as facial shape descriptors, where each surface normal component (x, y, or z) image is split into several local patches, all of them are further encoded by local normal patterns (LNP) on different scales. Liu and Sun (2015) introduced the local radial binary pattern (LRBP) to extract the normal angle feature of different regions, which significantly reduced the dimension of the feature vector, greatly alleviating the pressure of subsequent calculation. Soltanpour and Wu (2019) adopted the local derivative pattern (LDP) for three normal maps on different scales. An extreme learning machine-based autoencoder is used for dimension reduction.

4.3.2. Other methods

Statistical features. Chua et al. (2000) extracted the point signature (PS) on the rigid regions of the face, which selected a curve in the neighborhood of the keypoints as the local feature. Xu et al. (2006b) concatenated the depth features, cosine signature, and 0th–2th order moments of cosine signature features as an intrinsic features vector. The well-known machine learning algorithm, AdaBoost, is used to select the most efficient features in FR. 3D signature, a vector representation for 3D face shape, is proposed by Boehnen et al. (2009), which is composed of fixed surface points in a face-centered coordinate system. Inspired by the former work (Lin et al., 2005), a 2D and 3D geometrically invariant features, as is referred to summation invariants, is proposed by Lin et al. (2006a), which extracted both profile curves and surface patches from a rectangular region around the nose for recognition. An extension of this work is proposed in (2007), where summation invariants are extracted and weighted fused from multiple facial subregions.

Geometric features. An adaptive rigid multi-region selection (ARMS) algorithm is proposed by Chang et al. (2005b), which automatically finds the rigid regions in the high curvature area on the face without manually selecting landmarks. Actually, three nose regions with different shapes are selected and fused as the final result. Ming (2014) employed curvature information to extract the facial rigid areas, then an orthogonal spectral regression (OSR) algorithm is proposed to extract the high discriminant features on the manifold structure for recognition.

The local keypoint-based multiple triangle statistics (KMTS), an extension framework of (2013), is proposed by Lei et al. (2016). Each local surface is represented by multiple spatial triangles, and four types of geometrical features are calculated and concatenated to form the feature descriptor: e.g. the angle between the two lines connecting two vertices, the radius of the circle circumscribed to the triangle connecting the two vertices and the keypoint, the distance between the two vertices, the angle between the z-axis and the line connecting two vertices. Besides, a two-phase weighted collaborative representation classification framework is proposed to address the single sample problem. Experiments show this framework is robust to expression, pose variations and occlusion.

Covariance matrix descriptors. Besides encoding and fusing different types of features and modalities, the covariance matrix descriptors capture the correlation of these properties, eliminating the influence of some noise to some extent. Tabia et al. (2014) used the covariance matrices of the descriptors rather than the descriptors themselves to analyze the 3D shape. Hariri et al. (2016a) added other covariance descriptors into this representation (Tabia et al., 2014). Since covariance matrices are the elements of a Lie group based on a Riemannian structure, the geodesic distance in manifold space is converted to equivalent log-determinant distance or affine-invariant distance for matching. An extension of this work has been proposed (Hariri et al., 2016b), where three different levels of covariance descriptors are constructed around each feature point. The three description levels start from a small region to a bigger overlapped region so that fine local features and sufficient holistic properties can be captured at the same time. Križaj et al. (2013) used region covariance descriptors (RCMs) and Gaussian mixture models (GMMs) for FR. Specifically, the

Table 7
Local surface-based methods.

Reference	Main algorithm
Huang et al. (2006)	Global statistics of geometrical features and local 3D LBP histograms features
Huang et al. (2012)	Multiscale extended LBP
Lv et al. (2015)	Region-based extended LBP
Werghi et al. (2015b)	Mesh-LBP
Tang et al. (2013)	LBP applied to facial depth and normal maps
Li et al. (2014).	MSMC-LNP
Liu and Sun (2015)	Local radial binary pattern (LRBP)
Soltanpour and Wu (2019)	Local derivative pattern (LDP)
Tang et al. (2016)	Mesh-VHLBP
Chua et al. (2000)	Point Signature
Xu et al. (2006b).	Cosine Signature
Boehnen et al. (2009)	3D Signature
Lin et al. (2006a)	Summation Invariants
Ming (2014)	Orthogonal spectral regression (OSR)
Chang et al. (2005b)	Adaptive rigid multi-region selection (ARMS)
Lei et al. (2016)	Keypoint-based multiple triangle statistics (KMTS)
Hariri et al. (2016a)	Covariance descriptors
Hariri et al. (2016b)	Hierarchical covariance descriptors
Križaj et al. (2013)	Region covariance descriptors (RCMs)

3D facial data is mapped into a composite representation, just stacked one after another, from which a set of RCMs are extracted to model faces via GMMs.

Table 7 summarizes our surveyed local surface-based methods, where the descriptors are divided into LBP descriptors, geometric features, and statistical features neighboring the keypoints. Since the covariance matrices of descriptors are capable of capturing local geometric attributes and the correlation of these attributes, researchers commence to design new covariance matrices of the descriptors rather than use the descriptors themselves for FR.

5. Deep learning-based methods

Deep learning has received increasing interests in FR recently. Within this method, the low-level features are extracted to construct high-level essential features of the face through neural networks with multi hidden layers and massive training data. This section summarizes the main existing deep learning-based methods for 3DFR and groups them into three different categories: methods based on converting 3D to 2D data, designing novel deep learning networks, and 3D face reconstruction.

5.1. Methods based on converting 3D to 2D data

In the early time, methods based on converting 3D to 2D mainly concentrated on taking the depth image as the input of neural network for recognition. CNN is the most popular. Lee et al. (2016) employed transfer learning to fine tune CNN with an extremely small number of depth images, where CNN has been well trained on massive 2D face images beforehand. In addition, a 3D face expression augmentation technique is proposed to solve the lack of training data, which synthesizes different facial expressions through a single 3D face scan. Transfer learning has also been employed by Kim et al. (2017), where CNN is pre-trained with low-quality RGB images. Zhang et al. (2018c) first extracted the depth and color information from the RGB-D images as a joint feature, together with RGB-specific and depth-specific features fed into the branches of CNN for feature learning, both for the complementary features from multiple modalities and common features between different modalities.

Apart from the depth image, another 2D facial attribute map is described by Li et al. (2017a), where the 3D coordinators are projected into 2D plane to estimate three images of facial surface normal components, all of them are fed into a pre-trained VGGNet to extract deep features. An enhanced version of this work is presented by Li et al. (2017b), which uses six types of 2D facial attribute maps to represent the 3D face scan, e.g. geometry map, three normal maps, curvature map, and texture map, and all of them are jointly input into the branch of deep fusion CNN for feature extraction and fusion.

Feature learning based on combining different facial parts, texture and depth features performs better than only use whole face or local face regions. Jan et al. (2018) extracted four key facial parts (e.g. eyebrows, eyes, mouth, and nose) on a textured 3D face scan, all of which are resized and propagated individually through a pre-trained VGGNet. Cai et al. (2019) extracted complementary features from four facial patches with certain overlapping areas, including the whole face, upper half face, small upper half face, and separate nose area, all of which are input into ResNet for feature extraction and fusion.

There are many other methods designed to improve the performance of deep learning. The traditional loss function separates the between-class features and aggregates the intra-class features compulsorily, without considering the related attributes of the underlying faces. An attribute-aware loss function is proposed by Jiang et al. (2020), which incorporates some facial attributes (e.g. age, gender, and ethnicity) into training process and applies the attribute proximity to regularize the feature mapping. Mu et al. (2019) designed a lightweight and efficient CNN to recognize low-quality 3D faces, which uses a multi-scale feature fusion (MSFF) module to fuse features extracted from different convolution layers and sets corresponding weights via a spatial attention vectorization (SAV) module. In addition, a data processing system, including point-cloud recovery, surface refinement, and data enhancement, is proposed to deal with low-quality 3D faces to generate finer and bigger training data.

5.2. Methods based on designing new deep network

Recently, 3D voxel has been used to represent a geometric 3D shape. Wu et al. (2015) represented the 3D shape as a binary tensor defined on a 3D voxel grid, where 0 and 1 indicate the voxel is outside or inside the mesh surface respectively. 3D voxel has also been used for 3D shape recognition (Qi et al., 2016), which used volumetric CNN to extract features. 3D voxel is first employed to represent 3D face by Jackson et al. (2017), which used volumetric CNN to reconstruct 3D face from a single 2D face image, omitting the complex alignment and fitting process. Zhang et al. (2018b) proposed an end-to-end 3D facial landmark location method, called joint voxel and coordinate regression (JVCR). Specifically, a voxel regression subnetwork consisting of hourglass modules is adopted to estimate the volumetric representation from coarse to fine, followed by a coordinate regression subnetwork to regress the coordinate vector of the face shape.

As 3D voxel grids render data unnecessarily voluminous, a novel type of neural network is designed to directly input point cloud for feature extraction, named PointNet (Charles et al., 2017), which is irrelevant to the permutation of the point cloud. Tan et al. (2019) proposed an extension framework of PointNet, called deep registration network (DRNet), which works with ordered 3D point coordinates. The DRNet constructs the dense point cloud from 6 frames of sparse point clouds by registering and fusing, further converted to a depth map and fed into face recognition network (FRNet) for recognition. Trimech et al. (2020) extracted point-based representations by using various sampling strategies. The first is a set of 70 manual landmarks containing 68 annotated facial fiducial points and 2 automatically extracted cheek points, the second is a set of 204 points sampled from a level curve-based representation, and the third is a set of 300 SIFT keypoints, all of which are fed into PointNet for facial expression classification and recognition.

A deep CNN model, called FR3DNet, is first designed specifically for 3DFR (Gilani and Mian, 2018). The 3D point cloud is used to generate a three channels image, including depth image, azimuth, and elevation angles, and all of them are fed into CNN for deep feature extraction, in which every convolutional layer is followed by a rectifier layer.

Table 8
Deep learning-based methods.

Reference	Main algorithm
Lee et al. (2016)	Transfer learning
Kim et al. (2017)	Transfer learning
Zhang et al. (2018c)	Extract complementary and common features from depth and color image
Li et al. (2017a)	Estimate three images of facial surface normal components
Li et al. (2017b)	Use six types of 2D facial attribute maps to represent 3D face
Jan et al. (2018)	Four key facial parts
Cai et al. (2019)	Extract complementary features from four overlapping regions
Jiang et al. (2020)	Attribute-aware loss function
Mu et al. (2019)	Led3D: MSFF and SAV
Wu et al. (2015)	3D voxel used for 3D shape
Qi et al. (2016)	3D voxel and volumetric CNN
Jackson et al. (2017)	3D voxel used for 3D face
Zhang et al. (2018b)	Joint voxel and coordinate regression (JVCR)
Charles et al. (2017)	PointNet
Tan et al. (2019)	Deep registration network (DRNet)
Trimech et al. (2020)	Three sampling strategies based on point representations
Gilani and Mian (2018).	FR3DNet
Gilani et al. (2017)	Region-based 3D deformation model (R3DM)
Hsu et al. (2018)	Fast-hierarchical model (FHM)
AbdAlmageed et al. (2016)	Pose-specific CNNs
Kittler et al. (2018)	Each mesh vertex of 3DMM represented as 3D coordinators and associated R, G, B values
Dou et al. (2017)	UH-E2FAR: AFM+BFM
Feng et al. (2018)	Position map regression network (PRNet)
Liu et al. (2018b)	An encoder–decoder network
Dou and Kakadiaris (2018)	Deep recurrent 3D face reconstruction (DRFAR)
Sánchez-Escobedo et al. (2016)	Use 2D occluding contours to regress 3D vertices
Rara et al. (2010)	3D shape model and albedo model
Rara et al. (2011)	Use sparse 2D feature points to regress 3D model
Dou et al. (2014)	2D to 3D sparse landmarks and 3D sparse to dense shape
Tian et al. (2018)	Use a cascade of linear regressors to estimate 3D face shapes
Liu et al. (2020b)	Two cascaded regressors, one for updating 2D landmarks and the other for estimate 3D shape

Additional factors, e.g. facial subtle variations, camera viewpoint, and self-occlusion, are introduced to the limited 3D face scans to generate training datasets, while the existing 3D face datasets are merged as test datasets, thus obtaining the largest known 3D face dataset.

5.3. Methods based on 3D face reconstruction

Deep learning is also used together with 3D face reconstruction for recognition. Some methods used deep learning to detect facial landmarks to generate dense deformation model or learn features from 3DMM. Gilani et al. (2017) proposed a pre-trained deep network to detect facial landmarks, which are used to segment the 3D face into Voronoi regions by evolving horizontal iso-geodesic curves. A region-based 3D deformation model (R3DM) is then generated by establishing dense correspondence of every facial region. A fast-hierarchical model (FHM) is proposed to locate cross-pose facial landmarks (Hsu et al., 2018), then each 2D face is segmented and reconstructed separately. Two kinds of features, handcrafted features and learning features, are extracted from different facial components for recognition. AbdAlmageed et al. (2016) used 3D rendering to generate multiple face poses, all of which are fed into pose-specific CNNs for feature extraction. A 3DMM is constructed by Kittler et al. (2018), in which each mesh vertex represents the 3D coordinators and associated R, G, B values of the texture image. A conformal mapping is proposed to convert the 3D mesh to a 2D image, with being input into DCNN for learning and training.

As the fitting optimization of face alignment inevitably results in a significant computational cost, there has been a recent trend to explore end-to-end 3D face reconstruction bypassing the construction and fitting process. Jackson et al. (2017) used volumetric CNN regression to directly reconstruct 3D model from a single 2D face image. Three architectures of volumetric regression network (VRN) are proposed as follows: one is VRN, which accepts an RGB image as input and directly regresses a 3D volume. The second is VRN-Guided, which adds 2D projection of the 3D landmarks as a guide and regresses a 3D volume. The third is VRN-Multitask, which regresses both the 3D facial volume and a set of sparse facial landmarks. The best performing VRN is the one guided by detected landmarks (VRN-Guided). Feng et al. (2018) proposed a position map regression network (PRNet) to reconstruct 3D face and provide dense alignment simultaneously.

Dou et al. (2017) proposed a DNN-based approach for end-to-end 3D face reconstruction from a single 2D image. Two facial models, namely BFM model (Paysan et al., 2009) and AFM model (Kakadiaris et al., 2007), are used to construct the neutral face model and expressive face model via a multi-task loss function. With the fusion-CNN, the identity and expression features are extracted and fused to predict 3D expressive facial shapes. An encoder–decoder network is designed to disentangle identity and non-identity features during reconstruction (Liu et al., 2018b). Experiments demonstrate the disentangled identity features perform better for improving recognition accuracy. Compared with single-view 3D face reconstruction, multiple images of the same subject contain more facial information. Dou and Kakadiaris (2018) proposed a deep recurrent 3D face reconstruction (DRFAR) approach for multi-view 3D face reconstruction. The DCNN disentangles the facial identity and expression features of every single image, while the RNN fuses the identity-related features to predict the facial identity parameter as the final result.

Some machine learning methods considered 3D face reconstruction as a regression problem, either a single regressor or a cascade of regressors. Sánchez-Escobedo et al. (2016) trained a regression matrix to estimate 3D face surface from a set of 2D contours of input image. Two regression models, the 3D shape model and the albedo model, are constructed to overcome the effect of arbitrary illumination (Rara et al., 2010). An extension of this framework is designed to achieve low computational cost, which only needs sparse 2D feature points to regress 3D model (Rara et al., 2011). However, with the assumption of frontal poses and a Lambertian reflection model, this method limits the reconstruction accuracy in unconstrained conditions. Dou et al. (2014) proposed a two-fold coupled structure learning to reconstruct the monocular face shape. The 3D sparse landmarks coefficient is regressed by learning the relationship from 2D to 3D sparse landmarks, then a coupled dictionary learned on 3D sparse and dense shape is applied to reconstruct dense 3D face shape.

Tian et al. (2018) used a cascade of linear regressors rather than a single regressor as stated above to estimate 3D face shape, which can better extract complementary information of face images with various poses and expressions. Cascaded regression has also been used in (Liu et al., 2017, 2020a,b), different from (Tian et al., 2018), 2D landmark locations are estimated together with 3D shape for reconstruction. Specifically, the 2D landmarks regressor updated 2D landmarks while the shape regressor updated the 3D shape according to the landmarks update. Besides, the expression-related deformation has also been updated during updating 3D shapes, which can explicitly deal with expression deformation. The similarity scores of 2D images and reconstructed 3D models are separately calculated and fused in a weighted scheme.

Table 8 summarizes the existing deep feature-based methods. Most of these surveyed studies focus on converting 3D face to 2D attribute maps, as opposed to designing new deep network for 3DFR. And most of deep network architectures (e.g. AE, GAN, and RNN) mainly focus on 2D but rarely related in 3D, therefore, actively carrying out research in this area will be of great significance. In addition, end-to-end has become the mainstream in 3D face reconstruction.

6. Multimodal fusion-based methods

These methods use the fusion of 3D data and 2D texture information for FR, outperforming that employ only 3D or texture attributes, which can be grouped into three categories: signal level-based, feature level-based, and decision level-based fusion. At present, the majority of these methods adopt decision level-based fusion, and the other two are used relatively less.

6.1. Signal level-based

Signal level-based fusion is described as the fusion of 2D and 3D data in the initial stage of FR. Early-stage fusion is known to outperform later stage fusion, as early fusion data contains richer information about the input biometrics than the compressed features or matching scores. Papatheodorou and Rueckert (2004) combined 3D coordinates with associated 2D gray value as novel 4D face data for ICP registration. BenAbdelkader and Griffin (2005) concatenated the pixels of depth map and texture map, then a classifier for three different inputs, depth map, texture map, and combined depth and texture maps, is set up to calculate the similarity scores, all of which are summed at decision level. Experiments show that better recognition performance is achieved by using both depth and texture maps. PCA-based image recombination for multimodal 2D+3D FR is proposed by Kusuma and Chua (2011). Two types of recombined data are obtained as follows: the predominant 2D image with additional structure-cues and the predominant 3D image with additional texture-cues, then calculating and fusing the matching scores at decision level. Jiang et al. (2020) transformed 3D point cloud coordinates into three-channel images, with the value of R, G, B being extracted from color images to form six-channel images, all of which are further fed into ResNet for feature extraction.

6.2. Feature level-based

Feature level-based fusion usually extracts distinct features from 2D and 3D data, all of which are further fused as the feature vector for recognition. Mian et al. (2007) calculated spherical face representation as the 3D face descriptor and selected SIFTs as 2D local descriptors, both of them are combined as a rejection classifier to eliminate a large number of candidates' faces at the early stage. Xu et al. (2009) used Gabor filters to extract local features from depth and intensity images respectively, then employed LDA and AdaBoost learning to select the most effective features. Werghi et al. (2015a) introduced Mesh-LBP to extract shape and texture features on a triangular-mesh manifold, which are concatenated for recognition according to the distance of features.

6.3. Decision level-based

Decision level-based fusion uses two or more face classifiers to calculate the similarity scores of 2D and 3D data, then different fusion strategies are applied to fuse the matching scores.

Chang et al. (2005c) used PCA-based methods for depth image and color image and fused the matching scores with a weighted scheme. The embedded hidden Markov model (EHMM) is applied to depth and color images for face classification (Tsalakanidou et al., 2005). Lu and Jain (2005) used the modified ICP to match rigid variations, with the thin plate spline model being applied estimate non-rigid deformation. The matching distances are combined with the deformation classification results as the final decision. Lu et al. (2006) first used a set of 2.5D face scans at different views to construct 3D face model, further synthesizing 2D images at different poses and illumination. Then, the synthesized face images are used in LDA, while 3D face surface matching is based on a modified ICP. Finally, the weighted sum rule is applied to fuse the two matching scores.

Table 9
Multimodal fusion-based methods.

Reference	Main algorithm
BenAbdelkader and Griffin (2005)	Pixels of depth and texture map
Papatheodorou and Rueckert (2004)	3D coordinators and associated 2D gray value for ICP
Kusuma and Chua (2011)	PCA-based image recombination
Jiang et al. (2020)	3D coordinates and RGB value
Mian et al. (2007)	3D face descriptor: SFR 2D images: local SIFT
Xu et al. (2009)	Depth and intensity image: Gabor filters
Werghi et al. (2015a)	Triangular-mesh manifold: LBP
Chang et al. (2005c)	Depth and color image: PCA
Tsalakanidou et al. (2005)	Depth and color image: Embedded hidden Markov model (EHMM)
Lu and Jain (2005)	ICP matching distances and deformation classification results
Lu and Jain (2006)	2D synthesized face: LDA 3D face: ICP
Zhang and Wang (2011)	Detect scale space extrema on shape index images and texture images
Goswami et al. (2014)	RGB: entropy and saliency Depth: geometric facial attributes
Li et al. (2016a)	RGB and XYZ normal maps: multi-channel discriminant transform
Liu et al. (2020b)	2D images: DNN 3D models: ICP

Zhang and Wang (2011) detected scale-space extrema on shape index images and texture images respectively, then multi-order geometric features, including the keypoints features, the distance between points, the area and inner angles of the triangles formed by points triples, are extracted and fused at the decision level. Goswami et al. (2014) extracted texture features from RGB image by using methods based on entropy and saliency, with the geometric facial attributes being extracted from the depth image. The matching scores of both the attributes and features are fused to recognize face. Li et al. (2016a) transformed RGB maps and XYZ normal maps of 3D face into discriminative representation by using the multi-channel discriminant transform, both of them are then coded as a sparse combination with dynamic weight. A joint face alignment and 3D face reconstruction is proposed by Liu et al. (2020b). Specifically, the 3D face model is reconstructed from the 2D images, then DNN is used to obtain the similarity scores of 2D images, while ICP is used to match the reconstructed normalized 3D face models. Finally, the two scores are weighted and fused to get the recognition result.

Table 9 summarizes our surveyed multimodal fusion-based methods. The majority of multimodal fusion outperforms using single modal data, and the recognition accuracy of early-stage fusion is better than that of latter stage, however, the lower complexity seems to justify the employment of single modal or fusion at decision level. Moreover, the descriptors employed to represent face vary with the type of data. Studies that make use of 3D data usually extract geometric descriptors, e.g. 3D coordinate, mean and Gaussian curvatures, Euclidean, and geodesic distances, while LBP, SIFT descriptors are applied to deal with 2D texture information.

7. Research for 3DFR under unconstrained conditions

3D faces compared with 2D images are more robust in terms of aging, making up, occlusion, and variant PIE. However, extreme external conditions still undermine the recognition accuracy, which is a prominent challenge in FR. This section summarizes four common non-specific conditions of 3DFR: expression, pose, illumination variations, self-occlusion and spoofing attack.

7.1. Expression invariant

The captured 3D face is inevitable to deform non-rigidly with expression variations, which greatly undermines the performance of algorithms. Hence, expression invariant 3DFR has emerged as an active research area. This section surveys and explains the main existing expression invariant methods and group them into two different categories: local features and global features-based, and the latter can be fine divided into morphable model-based and facial surface-based methods.

7.1.1. Local features-based

Expression invariant methods based on local features only select the rigid regions which are insensitive to facial expressions for recognition. The most frequently used rigid regions are the nose and eye-forehead areas. Zhong et al. (2007) used Gabor filters to extract intrinsic features of the rigid region above the mouth. Cook et al. (2007b) employed Gaussian mixture model to weigh different facial regions. Actually, the model places a higher value on the nasal region, while reducing the contribution of the lips and cheeks under expression variations. Emambakhsh et al. (2013) detected 16 landmarks in the nose region and constructed 75 curves by connecting pairs of landmarks, of which 28 most consistent curves over facial expressions are selected for nose recognition. An extension of this framework is proposed in (2017), which locates 7 keypoints around the nose region and extracts histogram features based on the surface normal by using Gabor-wavelet filters.

7.1.2. Global features-based

Morphable model. The morphable models vary their shape in accordance with the target face to reconstruct 3D face model. In the following recognition, apart from directly using the morphable parameters, each of the reconstructed 3D models may be applied to synthesize 2D faces with different poses, expressions, and illumination. Jiang et al. (2005) reconstructed 3D face shape from a single frontal face image with neutral expression and normal illumination, then employed the face texture extracted from 3D model to synthesize virtual faces with different PIE. A similar method is employed for pose invariant FR (Zhao et al., 2017). Prabhu et al. (2011) applied 3D generic elastic model to generate a 3D model, which is further rendered at different poses to obtain a set of 2D face images. Chang et al. (2018) used CNN to directly extract 3D facial expression coefficients from image intensities, further establishing 3D expression model without detecting facial landmarks.

In addition, 3D face model can be applied to modify the original 3D face to the neutral frontal face. Lu and Jain (2008) proposed a hierarchical geodesic-based resampling approach to extract landmarks for modeling facial deformations in the control group. The learned deformation is then transferred to 3D neutral model to synthesize 3D face model with variant poses and expressions, all of which are further combined to construct the deformable model for the fitting process. Zhu et al. (2015) used 3DMM to estimate pose, expression and identity parameters from 2D images. The high-fidelity pose and expression normalization (HPEN) is introduced to normalize the pose and expressions, with Poisson Editing being employed to fill the invisible region, thus a smooth frontal face with neutral expression is obtained. Tang et al. (2019) exploited the 3DMM to generate faces of desired poses. The representation-learning Wasserstein-GAN with three component networks is proposed to disentangle identity, pose, and expressions, further constructing the training set with uniform facial pose distribution.

Since the synthetic face and real face are from different domains, the model only trained by synthetic faces might not be well extended to real faces. Feng et al. (2015) proposed a cascaded collaborative regression to detect facial landmarks by using a mixture of synthetic and real faces with dynamic weight. The synthetic faces are used heavily in the beginning while more real faces are incorporated to model finer detail as the training proceeds. An et al. (2017) used deep transfer network to reduce bias between the 2D real face images and 3D synthesized face images, where the maximum mean discrepancy of the shared feature extraction layers and the discrimination layers is calculated to further optimize the transfer network. The results show that jointly using synthesized images and real images performs better than using either dataset. An extension of this framework is proposed in (2018), which chooses both AlexNet and Inception-ResNet-V1 as the benchmark models.

Facial surface. Bronstein et al. (2005) assumed that the facial surface deformation caused by facial expressions can be modeled as isometric transformation, that is, the geodesic distance between any two points on the surface remains approximately invariant across different expressions. On this basis, multidimensional scaling is applied to the face surface, further converting the geodesic distance into equivalent Euclidean distance. An extension of this framework is proposed in (2006), where a generalized multidimensional scaling is employed to handle partially missing data caused by occlusion. Based on this model, Li et al. (2008) constructed the radial geodesic distance image to generate geodesic level curves centered at nose tip, then deformation invariant image is constructed by evenly sampling points from the selected curves. A guidance-based constraint deformation (GCD) is proposed by Wang et al. (2007), which uses the Poisson equation to complete the deformation from expressive to neutral face model, with rigid constraints being employed to ensure the inter-class differences. The main limitation of these methods is the introduced expression processing model is insufficient to approximate faces. However, there may be no such model which is fully adapted to the face with arbitrary expression.

7.2. Pose invariant

The majority of 3DFR need to convert the original face to a frontal face since the face scan has difficulty in being matched with the present face model when the head pose deviates greatly from the front-face pose. In this section, we provide a comprehensive review of pose invariant methods, which can be classified into two different categories: pose correction-based and synthetic face-based methods.

7.2.1. Pose correction-based

Most of these methods perform pose correction by locating the rigid landmarks of the face, which is also the necessary step in the preprocessing. The commonly used facial landmarks, apart from the nose tip, include the inner or outer eye corner, the mouth corner and so on.

Xu et al. (2006a) presented a hierarchical filtering scheme to locate nose tip, then the nose ridge is estimated by included angle curve (ICA) based on the detected nose tip. Lin et al. (2006b) introduced a coupled 2D and 3D feature extraction approach to locate the eye sockets, while the nose tip is considered as the extreme vertex along with the normal directions of eye sockets. Lu and Jain (2006) proposed a feature extractor to estimate the nose tip and pose angle, with 3D feature location model being used to extract eye and mouth corners. Wei et al. (2007) used the surface normal difference (SND) algorithm and shape index estimation to locate the nose tip and nose bridge.

Mian et al. (2007) introduced a heuristic method for nose tip detection on the basis of geometric analysis of the nose ridge contour. Romero-Huertas and Pears (2008) presented a graph matching method to locate the inner eye corners and nose tip, where the "distance to local plane" notion is introduced to eliminate unlikely candidates and then selected the best combination of landmarks for recognition. Faltemier et al. (2008) proposed a technique called rotated profile signatures (RPS) for nose detection and pose categorization. With the rotation of 3D face over a 180 interval in 5 increments, the points on the rightmost facial edge at each step are extracted and matched with a set of nose models, while the minimum similarity score indicates the

correct nose location. Peng et al. (2011) proposed a training-free nose tip detection method, which can handle both frontal and non-frontal poses at a relatively fast speed. 2D Left-and Right-Most face profiles are generated to detect nose tip candidates, then the cardinal point fitness and spike fitness are calculated to identify the nose tip.

Passalis et al. (2011) employed an automatic landmark detector to estimate pose and detect occluded areas, and applied the facial symmetry to handle the missing data when fitting AFM to the target face. An extension of this framework is proposed in (2013), 3D local shape descriptors, including shape index and spin images, are used to extract candidate landmarks, all of which are matched with a set of facial anatomical landmarks (e.g. the eye inner/outer corner, nose tip, mouth corner, and chin tip) for face identification. However, these pose correction methods based on landmark detection are not robust under significant pose variations.

7.2.2. Virtual face-based

In virtual face-based methods, a statistic model learned beforehand is applied to either make the input faces achieve "frontalization" or render a set of virtual faces at the desired pose, making two faces at different poses matched at the same pose. Hassner et al. (2015) used an average 3D face model to approximate the shape of all input faces to achieve "frontalization". Kim et al. (2016) used 2D facial landmarks together with generic 3D face model to estimate head pose. Kan et al. (2014) proposed the stacked progressive auto-encoders (SPAE) to transform the non-frontal faces images to frontal ones. Specifically, each shallow auto-encoder is designed to convert the input face image at a large pose to a virtual view at a smaller pose, with those already closer to frontal keeping unchanged. In other words, SPAE gradually narrows down the pose variations layer by layer, which makes the output of the topmost layers can be regarded as a frontal face. Zhang et al. (2019) proposed a pose-weighted GAN to synthesize frontal faces by refining the pose code in the pose-weighted pixel loss to address the face in large pose. Besides, the HPEN is used to add the 3D structure constraints for self-occluded regions.

Apart from the most popular statistical model, many other deep learning methods are adopted to synthesize virtual faces at different poses. Zhang et al. (2018a) designed a GAN-based structure to synthesize facial images with different expressions and poses. Specifically, an encoder–decoder structure is served as a facial image converter. For the encoder, the input is a face image with arbitrary pose and expression, while the output is a synthetic facial image at target pose and expression. For the decoder, two discriminators are used to disentangle the attributes (e.g. expression and pose) from identity for further recognition. Deng et al. (2018) proposed the local and global adversarial DCNNs to learn the complete UV map, which is attached to the fitted 3D mesh to generate the virtual face with arbitrary poses. Chung et al. (2020) proposed a novel generative model to enhance UV map completion. Specifically, the incomplete UV extracted from in-the-wild images are improved by stacking two U-Nets based on ResNet architecture, and the multiple-level residual connections between two U-Nets are applied to reduce gradient vanish and promote feature fusion.

Contrary to the above methods, either learn to pose invariant features by training on massive data or align the original face to a single frontal face, some methods are proposed to directly handle pose variations. Liu et al. (2018a) proposed a multi-channel pose-aware CNN, which uses three sub-CNNs to extract features from different facial areas (e.g. the whole face, eyes, and mouth), further fed into pose-specific recognition branches to estimate the facial pose. Masi et al. (2019) proposed pose-aware models to process face images, which uses several pose-specific DCNN to tackle pose variations and all of these models have been pre-trained by multiple synthetic face poses.

7.3. Occlusion invariant

Face occlusion, caused by hair and hands or external objects, e.g. glasses, hats, and scarves, may lead to a failure of recognition algorithm. Hence, occlusion detection and restoration techniques seem particularly significant in 3DFR. On the other hand, Spoofing Attack, including printed images, replayed videos and 3D masks, still remains a serious problem. This section reviews the common used strategies in face occlusion invariant and anti-spoofing to 3D face mask. The former is grouped into facial curves-based, non-occluded facial regions-based methods, where the latter can be mainly categorized into appearance-based methods, motion-based methods, and remote photoplethysmography (rPPG).

7.3.1. Face occlusion

Facial curves. The radial curves, defined as open curves that start from the nose tip and end on the edge of the face, are the key facial curves in the presence of occlusion. Drira et al. (2010) used the radial curves derived from the nose tip to represent the 3D face. The discontinuous or too short curves are discarded while the remaining curves are employed to construct a Riemannian analysis framework. In the case of occlusion, the recursive-ICP is applied to detect each radial curve and remove the occlusion, while the statistical estimation on shape manifold of curves is used to recover the missing data. An extension of this framework is proposed by Gawali and Deshmukh (2014), the radial geodesic curves are extracted on 3D meshes, again, the elastic shape analysis is used to compare shapes of facial curves. The radial strings have also been used by Yu et al. (2016a,b), which uses dynamic programming to match the corresponding attributed string sets and AdaBoost algorithm to select the most discriminative radial strings for recognition. Moreover, the occluded parts are eliminated by a partial matching mechanism.

There are many other facial curves are adopted for FR in the presence of occlusion. Li and Da (2012) extracted the central profile in the nose region to form a rejection classifier, each of the remaining faces is segmented into six facial regions. The facial curves in these regions, which are the most sensitive to facial expression and hair occlusion, are extracted to evaluate facial deformation. Berretti et al. (2013b) combined SIFT with spatial clustering to detect keypoints on the depth image and then connected pairs of keypoints to get the facial curves, along which the depth value is modeled. The distances between facial curves weighted by their saliency are selected as the similarity measure. Smeets et al. (2013) directly extracted SIFT keypoints on 3D mesh, where the left–right symmetry of the human face allows comparing two facial surfaces with missing data.

Non-occluded facial regions. These methods select all non-occluded facial regions adaptively instead of using only the nose and eye-forehead regions or the whole facial surface for recognition. Alyuz et al. (2008, 2010, 2012a) and Alyüz et al. (2012b) divided the whole face surface into local components after coarse alignment, all of which are separately registered to relative Average Region Models (ARMs) by dense alignments. Moreover, the thresholding technique is employed to detect and reject the information from the occluded regions, where the thresholds is calculated on the non-occluded neutral face.

Based on this framework, the group proposed new approaches to deal with occlusion. The nose region is registered based on coarse-to-fine alignment, then the occluded areas are automatically detected via Average Face Model (AFM). The Gappy PCA is used to restore the full face after removing the non-facial parts (Alyuz et al., 2012a; Alyüz et al., 2012b). Alyuz et al. (2013) proposed an adaptively selected model to detect the non-occluded regions for alignment. For the classification problem, the masked projection is employed to handle the missing data after the occlusion has been removed. Alyuz et al. (2014) proposed two occlusion detectors. In the first case, Gaussian mixture models are applied to estimate the difference between the target face

and AFM, further determining whether each region is occluded or not. The second method incorporated neighboring pixel relations into the model, where each pixel weight is set by using fitness to pixel-wise and neighboring pixel-pairwise relations. Then the graph cut technique (2006) is applied to detect the occlusion.

The thresholding technique has also been used by Bagchi et al. (2014), which is committed to generalize the thresholding technique for the detection of real-world occlusion. The threshold-based method has also been used on RGB-D images to localize the occluded areas (Zohra et al., 2016). Assuming that the occluded regions are close to the camera than other regions, therefore, the connected component with higher intensity values is selected as the occluded region (Saxena et al., 2007).

Apart from the threshold-based method, the block-based method (Ganguly et al., 2015) is proposed for occlusion detection. Based on that obtrude sections have high depth density than any other facial surface, two blocks of different sizes are rolled over the whole facial surface in a row-major order to calculate the depth values, effectively identifying the outward and occluded facial regions. In (2012), a rejection strategy is embedded into ICP to eliminate the impacts of occlusion, which has been applied in (2006; 2005). A nose detection approach based on template matching of depth images is proposed by Liu et al. (2014). Three nose templates, called the Average Nose Model (ANM), are constructed for the whole nose, the left part of nose, and the right part of nose respectively. Hence, the nose can be accurately detected even in the presence of occlusion. Bellil et al. (2016) introduced Gappy wavelet neural network (GWNN) to cope with facial occlusion. Specifically, the Gappy wavelet transform is first applied to detect the face, where the occluded regions are defined as local face deformations by comparing the wavelet coefficients of the input face and generic face model. A multi library wavelet neural network is employed to restore occlusion.

7.3.2. 3D face mask

Face anti-spoofing is recently of great interest to some researchers, 3D face masks seem particularly challenging than printed images and replayed videos in FR system. In the recent years, studies on face anti-spoofing to 3D masks mainly contains the following three methods.

Appearance-based methods. In these methods, the face mask is distinguished from real face via micro appearance difference (color or texture). Maatta et al. (2011) used multi-scale LBP to analyze the texture of facial images and encoded micro-texture patterns into an enhanced feature histogram. Wen et al. (2015) extracted four different features (specular reflection, blurriness, chromatic moment, and color diversity) to form image distortion analysis feature vector for face spoof detection. Chen et al. (2020) proposed a multi-modal dynamics fusion network, which encoded texture and shape clues by a two-branch deep CNN model. Since their poor generalization ability, these methods are vulnerable to the variations in acquisition conditions.

Motion-based methods. These methods are usually used to detect facial image attacks or hard 3D face masks by capturing subconscious motion of organs and muscles, e.g. eye blink (Sun et al., 2007), mouth movement (Kollreider et al., 2007) and head rotation (Bao et al., 2009), while they may fail to handle soft 3D face masks which possesses vivid facial skin motions, e.g. latex (Agarwal et al., 2017) and silicone (Bhattacharjee et al., 2018).

Remote photoplethysmography (rPPG). The rPPG captures the pulse change and subtle facial skin color variation caused by heartbeat (Liu et al., 2016), overcoming the limitations of appearance and motion methods, while it is sensitive to modeling views and lighting conditions. Li et al. (2016b) extracted the global rPPG from the facial region and use the signal amplitude for 3D mask detection. Inspired by this model, a multi-channel rPPG corresponding feature with the global noise-aware template learning and verification framework is proposed by Liu et al. (2021). Yao et al. (2021) extracted multiple ROIs covering whole face, while the facial regions containing richer rPPG signals are emphasized with larger weights, making it robust enough to noise and object motion.

7.4. Illumination invariant

Unconstrained illumination results in significant variations in the images of faces and thus affects the performance of 3DFR. Model-based methods are the most popular, besides estimating illumination coefficients directly, unlighting and relighting are common used to handle illumination problem, where the former normalizes the face as standard illumination, and the latter aims to generate a set of face images under different illumination conditions for recognition.

Romdhani et al. (2006) introduced the 3DMM as a generative model to estimate shape and texture information from a single 2D image, and combined the 3D shape information and reflectance models to describe the effects of illumination. The spherical harmonic illumination representation is introduced for FR under arbitrary illuminations (Zhang and Samaras, 2006). Two methods are proposed in this paper, the first method constructs a 2D statistical model to estimate the spherical harmonic basic images, which can only recognize the 3D face with a fixed pose. The second method introduces 3DMM on the basis of spherical harmonic illumination representation, which can recover shape and texture information from images across both poses and illuminations. A 3D spherical harmonic basis morphable model is proposed in (2009), where the shape parameters, spherical harmonic basis parameters, and illumination coefficients are extracted to represent the target face under unknown pose and lighting. Besides, Markov random field is introduced to represent the statistical distribution and spatial coherence of facial texture features, effectively overcoming extreme lighting conditions and partially occlusion. The sphere harmonic illumination representation has also been employed by Ma et al. (2014), which adopted the sphere harmonic illumination model to the 3DMM fitting process. In (2019), spherical harmonic lighting coefficients are estimated together with 3DMM coefficients, while the face normal, face mask, face shading, and face albedo are then extracted from the reconstructed 3D model. New shading and relit images are rendered under random lighting conditions.

Apart from the spherical harmonic illumination representation, many other model fitting methods are introduced to separate illumination information from identity-relative features. Rara et al. (2011) only need a set of sparse 2D feature points of the input image to recover the 3D facial shape. Aldrian and Smith (2013) decoupled the geometric and photometric information of 3D face image formation process, and proposed two methods to recover facial texture and specular reflectance after 3D shape model fitting, including unconstrained illumination and specular invariant model fitting with the assumption of arbitrarily distributed but single, known illumination. The facial symmetry constraint is introduced to determine the proportion of albedo and shading during the fitting process (Hu et al., 2013). Wu and Deng (2016) employed 3DMM to generate multi-views and various illumination samples, all of which are applied to train a multi-task DNN for pose and illumination normalization.

Some approaches focus on the face illumination transfer problem, which changes the lighting of source face image to reference face image with preserving other attributes. This method is applied easily to illumination invariant FR by transferring a standard illumination to all the face images.

Han et al. (2008) first converted the target face image to a logarithm domain, on which applied a homomorphic wavelet filtering to estimate the illumination information for relighting. Chen et al. (2013) decomposed the images into color, large-scale, and detail layers through the weighted least squares according to the gradient values, then the largescale layer of input image is replaced by that of reference image which has been soothed via the adaptive edge-preserving filter in advance. Besides, a method of face illumination normalization based on similar reflectance prior is proposed to deal with non-normal lighting face. Jin et al. (2020) first estimated the surface model and the albedo image from the reconstructed 3D model, both of them are aligned and filled by detecting facial landmarks for light rendering and illumination

Table 10
Expression invariant methods.

Reference	Main algorithm
Zhong et al. (2007)	Extract Gabor intrinsic features of rigid region above the mouth
Cook et al. (2007b)	Employ Gaussian Mixture Model to weight facial regions
Emambakhsh et al. (2013)	Detect 16 landmarks in the nose region
Emambakhsh and Evans (2017)	Detect 7 keypoints around the nose region
Jiang et al. (2005)	Synthesize virtual faces with different PIE
Zhao et al. (2017).	Synthesize virtual faces with different poses
Prabhu et al. (2011)	3D Generic Elastic Model
Lu and Jain (2008)	Hierarchical geodesic-based resampling approach
Zhu et al. (2015)	High-fidelity Pose and Expression Normalization (HPEN)
Tang et al. (2019)	Representation-Learning Wasserstein-GAN (RL-GAN)
Feng et al. (2015)	Cascaded Collaborative Regression (CCR)
An et al. (2018)	Deep Transfer Network (DTN)
Bronstein et al. (2005)	Assume facial surface deformation caused by expressions as isometric transformation
Li et al. (2008)	Deformation invariant image
Wang et al. (2007)	Guidance-based Constraint Deformation (GCD)

Table 11
Pose invariant methods.

Reference	Main algorithm
Xu et al. (2006a)	Nose tip: hierarchical filtering Nose ridge: Included Angle Curve (ICA)
Lin et al. (2006b)	Detect nose tip according to the extreme vertex along the normal directions of eye sockets
Wei et al. (2007)	Use Surface Normal Difference (SND) and shape index estimation to detect nose tip and nose bridge
Mian et al. (2007)	Use geometric analysis of the nose ridge contour to detect nose tip
Romero-Huertas and Pears (2008)	Use graph matching to locate I-inner eye corners and nose tip
Peng et al. (2011)	A training-free nose tip detection method
Faltemier et al. (2008)	Use Rotated Profile Signatures (RPS) to detect nose and categorize pose
Lu and Jain (2006)	Use a feature extractor based on the directional maximum to estimate nose tip and pose angle
Passalis et al. (2011)	Use an automatic landmark detector to estimate pose and detect occluded areas
Perakis et al. (2013)	Use 3D local shape descriptors to extract candidate landmarks
Hassner et al. (2015)	Average 3D face model
Kim et al. (2016)	Combine 2D facial landmarks and generic 3D face model to estimate head pose
Kan et al. (2014)	Stacked Progressive Auto-Encoders (SPAE)
Zhang et al. (2019)	Pose-Weighted GAN
Zhang et al. (2018a)	Use GAN to synthesize facial images with different poses
Deng et al. (2018)	Use local and global adversarial DCNNs to learn complete UV map
Chung et al. (2020)	Use two stacked U-Nets to learn complete UV map
Liu et al. (2018a)	Multi-channel pose-aware CNN
Masi et al. (2019)	Pose-Aware Models (PAM)

Table 12
Occlusion invariant methods.

Reference	Main algorithm
Drira et al. (2010)	Radial curves
Gawali and Deshmukh (2014)	Radial geodesic curves on 3D meshes
Yu et al. (2016a,b)	Boosting radial strings
Li and Da (2012)	Use facial curves to form a rejection classifier and adaptively selects non-occluded regions
Berretti et al. (2013b)	Use facial curves connecting pairs of SIFT keypoints to model the depth value along curves
Smeets et al. (2013)	Use left–right symmetry of face to deal with missing data
Alyuz et al. (2008, 2012a), Alyüz et al. (2012b)	Average Face Model (AFM)
Alyuz et al. (2010)	Average Region Model (AvRM)
Alyuz et al. (2012a), Alyüz et al. (2012b)	Detect occluded areas via AFM and restore full face via Gappy PCA
Alyuz et al. (2013)	Select non-occluded regions adaptively and handle missing data via masked projection
Bagchi et al. (2014)	Generalize the thresholding technique to detect real-world occlusion
Zohra et al. (2016)	Thresholding technique applied to RGB-D images
Saxena et al. (2007)	Average Nose Model (ANM)
Ganguly et al. (2015)	Use the block-based method for occlusion detection
Alyuz et al. (2014)	Two occlusion detection approaches: statistical facial surface model and binary image segmentation
Liu et al. (2014)	A rejection strategy embedded into ICP
Bellil et al. (2016)	Gappy Wavelet Neural Network (GWNN)
Maatta et al. (2011)	Multi-scale LBP
Wen et al. (2015)	Image distortion analysis feature vector (specular reflection, blurriness, chromatic moment, and color diversity)
Chen et al. (2020)	Two-branch deep CNN model applied to encode texture and shape clues
Sun et al. (2007)	Eye blink
Kollreider et al. (2007)	Mouth movement
Bao et al. (2009)	Head rotation
Agarwal et al. (2017)	MLFP database using latex and paper masks latex
Bhattacharjee et al. (2018)	CS-MAD database using custom-made flexible silicone masks.
Liu et al. (2016)	Analyze heartbeat signal via rPPG
Li et al. (2016b)	Extract the global rPPG from the facial region
Liu et al. (2021)	Multi-channel rPPG with global noise-aware template learning
Yao et al. (2021)	Weighted Spatial–Temporal Representation

transfer, finally achieving the illumination swap between two faces. Xu et al. (2021) introduced GAN to address the instance-level face illumination transfer, with the illumination-inspired attention mechanism being conducted to better deal with salient lighting effects.

Tables 10–13 summarize our surveyed 3DFR under unconstrained conditions respectively. The expression invariant-based category considers local features and global features. The former only selects the rigid regions less affected by expression variations, while the latter contains two main methods: the one is to utilize morphable model to synthesize virtual faces at desired expression, and the other is to establish a unified expression model to transform the target face as a neutral face. The second category, pose invariant-based methods, adopts either the rigid landmarks or synthetic face for pose correction. Occlusion invariant-based methods, our third method, mainly extract facial curves and non-occluded facial regions for occlusion detection and restoration. As for face anti-spoofing, rPPG seems to be the mainstream than appearance-based or motion-based methods. The final category, illumination invariant-based methods, apart from the most-used morphable model, pre-defined face normal maps and reflectance-like image are used to handle illumination variations as well.

8. Discussion and challenges

8.1. Discussion

3DFR has received widespread attention, as it can effectively overcome the limitations of 2DFR in terms of variant PIE and occlusion. This paper reviews the common 3DFR methods covering from traditional to modern works over the past two decades, along with the challenges and future direction of potential research.

All surveyed 3DFR systems in this paper can be categorized as traditional and modern methods. Traditional methods extract distinctive compact features on the facial surface for matching, e.g. global features, local features, and hybrid features. The former is mainly based

Table 13
Illumination invariant methods.

Reference	Main algorithm
Zhang and Samaras (2006)	Spherical harmonic illumination representation
Wang et al. (2009)	Combine 3D spherical harmonic basis morphable model and Markov random field
Ma et al. (2014)	Sphere Harmonic Illumination Model (SHIM)
Romdhani et al. (2006)	Combine the 3D shape information and reflectance models
Aldrian and Smith (2013)	Decouple the geometric and photometric information to reconstruct textured 3D model
Hu et al. (2013)	Use the facial symmetry constraint to determine the proportion of albedo and shading clues
Rara et al. (2011)	Use a set of sparse 2D feature points to recover 3D facial shape
Wu and Deng (2016)	Use 3DMM to generate multi-views and various illumination samples
Le and Kakadiaris (2019)	Estimate coefficients of spherical harmonic lighting and 3DMM
Han et al. (2008)	Use homomorphic wavelet filtering to estimate the illumination information
Xu et al. (2021)	Use Relight GAN to address the instance-level face illumination transfer
Jin et al. (2020)	Use dense landmark and semantic parsing for face illumination transfer and swapping
Chen et al. (2013)	Use a single reference image by adaptive layer decomposition for face illumination manipulation

on subspace or 3DMM, while the latter are subdivided into keypoints, curves, and local surface-based methods. These methods generally map the 3D face into 2D representation, and part of them have to combine with intensity images for recognition. Some discriminative information of face may be lost during compression of facial features from 3D to 2D, leading to inactivity of the traditional methods todays.

On the contrary, the modern deep learning methods used to identify or rebuild 3D face have recently been further explored in this field. There is no need to define the ROIs, extract features and select them, just simply feeding the preprocessed 3D face into deep network to learn the proper features, with emphasis on designation of network architecture and loss function. However, the research community has worked on the conversion of 3D into 2D and then adopts deep learning essentially in the 2D domain, as opposed to designing new deep network specifically for 3DFR. Some recent works use deep learning for end-to-end 3D face reconstruction bypassing the complicated fitting process, and the restricted model is robust under different PIE. In addition, 3D reconstruction methods based on X-ray have been applied to mummies or artifacts, which may be the promising trends in FR.

In addition, this survey summarizes some of the existing challenges, including pose, expression, illumination variations and self-occlusion. Before feeding 3D face to an FR module, face anti-spoofing is an essential part to judge whether the face is live or spoofed, which avoids different types of attack to some extent. Besides selecting comparatively rigid features of the face, most of these methods used morphable model to synthesize virtual face at desired PIE to overcome the influence of degradation conditions. Nevertheless, 3DFR under unconstrained conditions is still a tremendous challenge in the future.

8.2. Challenges

In particular, no existing 3DFR methods are capable of handling all the challenges, to the best of our knowledge, the key technique challenges are as the following aspects.

3D face datasets. The traditional 3D face scanners are expensive and time-consuming, requiring the subject to maintain a fixed pose for a long time, which places restrictions on its promotion to a large scale. On the other hand, the emerging RGB-D cameras not only require specific preprocessing on the obtained images but also combine the depth and color images for recognition. All of these drawbacks, as well as unconstrained conditions limit the size and scale of 3D face datasets. How to rapidly obtain high accuracy 3D face data in unconstrained conditions with comparatively low acquisition cost and scale is the key issue we should focus on in the future.

The accuracy and efficiency of algorithm. Existing 3DFR methods concentrate on maintaining the algorithm performance, of which precision and efficiency are studied to a lesser extent. Most of these algorithms are too cumbersome to be directly applied to large-scale datasets for real-time recognition. How to improve the speed of the algorithm with the accuracy being ensured, appears to be considerable in 3DFR.

Multimodal face recognition. Even though most multimodal methods outperform those employ only 3D or texture attributes, single modal data is used comparatively more for the lower calculation costs and complexity. In addition, early-stage fusion surpasses later stage fusion, since it may take full advantage of the different modal information rather than the compressed features or matching scores, however, most of researches adopt fusion at decision level, so more in-depth exploration is required for signal or feature level.

3DFR under unconstrained conditions. The variant PIE, occlusion, identity bias are still the issues to be solved in 3DFR. Most of existing methods select the facial rigid regions or render new faces with different PIE for recognition. However, all of them inevitably omit part of facial discriminative information. We anticipate that with technological advances, more concise and effective algorithms will be offered to recognize 3D face under unconstrained conditions.

9. Conclusions

This survey paper summarizes the main 3DFR methods covering traditional and modern methods, as well as their benefits and limitations. In addition, some face recognition challenges (e.g. variant PIE, self-occlusion, and spoofing attack), are also addressed in this paper. Future work may include a comparative survey of novel deep learning techniques applied in 3DFR. Besides helping fresh researchers to select the most effective method for the right situation, we hope this survey may motivate experienced ones to dedicate more consideration and attention to this area.

CRediT authorship contribution statement

Menghan Li: Writing – original draft. **Bin Huang:** Writing – review & editing. **Guohui Tian:** Resources, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is Supported by National Natural Science Foundation of China (61773239, 61973187), and the Taishan Scholars Program of Shandong Province.

References

AbdAlmageed, W., Wu, Y., Rawls, S., Harel, S., Hassner, T., Masi, I., Choi, J., Lekust, J., Kim, J., Natarajan, P., Nevatia, R., Medioni, G., 2016. Face recognition using deep multi-pose representations. In: 2016 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 1–9.

Agarwal, A., Yadav, D., Kohli, N., Singh, R., Vatsa, M., Noore, A., 2017. Face presentation attack with latex masks in multispectral videos. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. CVPRW, pp. 275–283.

Al-Osaimi, F.R., Bennamoun, M., Mian, A., 2008. Integration of local and global geometrical cues for 3D face recognition. Pattern Recognit. 41, 1030–1040.

Al-Osaimi, F., Bennamoun, M., Mian, A., 2009. An expression deformation approach to non-rigid 3D face recognition. Int. J. Comput. Vision 81, 302–316.

Aldrian, O., Smith, W.A.P., 2013. Inverse rendering of faces with a 3D morphable model. IEEE Trans. Pattern Anal. 35, 1080–1093.
Alexandre, G.R., Soares, J.M., Pereira Thé, .G.A., 2020. Systematic review of 3D facial expression recognition methods. Pattern Recognit. 100, 107108.
Alyuz, N., Gokberk, B., Akarun, L., 2008. A 3D face recognition system for expression and occlusion invariance. In: 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems. pp. 1–7.
Alyuz, N., Gokberk, B., Akarun, L., 2010. Regional registration for expression resistant 3-D face recognition. IEEE T Inf. Foren. Sec. 5, 425–440.
Alyuz, N., Gokberk, B., Akarun, L., 2012a. Adaptive registration for occlusion robust 3D face recognition. In: Computer Vision ECCV 2012. Workshops and Demonstrations. Berlin, Heidelberg, pp. 557–566.
Alyuz, N., Gokberk, B., Akarun, L., 2013. 3-D face recognition under occlusion using masked projection. IEEE Trans. Inf. Foren. Sec. 8, 789–802.
Alyuz, N., Gokberk, B., Akarun, L., 2014. Detection of realistic facial occlusions for robust 3D face recognition. In: 2014 22nd International Conference on Pattern Recognition. pp. 375–380.
Alyüz, N., Gökberk, B., Spreeuwers, L., Veldhuis, R., Akarun, L., 2012b. Robust 3D face recognition in the presence of realistic occlusions. In: 2012 5th IAPR International Conference on Biometrics. ICB, pp. 111–118.
Amberg, B., Knothe, R., Vetter, T., 2008. Expression invariant 3D face recognition with a morphable model. In: 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. pp. 1–6.
An, Z., Deng, W., Hu, J., 2017. Deep transfer network for face recognition using 3D synthesized face. In: 2017 IEEE Visual Communications and Image Processing. VCIP, pp. 1–4.
An, Z., Deng, W., Yuan, T., Hu, J., 2018. Deep transfer network with 3D morphable models for face recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 416–422.
Bagchi, P., Bhattacharjee, D., Nasipuri, M., 2014. Robust 3D face recognition in presence of pose and partial occlusions or missing parts. Int. J. Found. Comput. Sci. Technol. 4, 21–35.
Ballihi, L., Amor, B.Ben., Daoudi, M., Srivastava, A., Aboutajdine, D., 2012. Boosting 3-D-geometric features for efficient face recognition and gender classification. IEEE Trans. Inf. Foren. Sec. 7, 1766–1779.
Bao, W., Li, H., Li, N., Jiang, W., 2009. A liveness detection method for face recognition based on optical flow field. In: 2009 International Conference on Image Analysis and Signal Processing. pp. 233–236.
Bellil, W., Brahim, H., Amar, C.B., 2016. Gappy wavelet neural network for 3D occluded faces: detection and recognition. Multimed. Tools Appl. 75, 365–380.
BenAbdelkader, C., Griffin, P.A., 2005. Comparing and combining depth and texture cues for face recognition. Image Vis. Comput. 23, 339–352.
Berretti, S., Del Bimbo, A., Pala, P, 2010. 3D face recognition using isogeodesic stripes. IEEE Trans. Pattern Anal. 32, 2162–2177.
Berretti, S., Del Bimbo, A., Pala, P, 2011. 3D partial face matching using local shape descriptors. In: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding. pp. 65–71.
Berretti, S., Del Bimbo, A., Pala, P, 2013b. Sparse matching of salient facial curves for recognition of 3-D faces with missing parts. IEEE Trans. Inf. Foren. Sec. 8, 374–389.
Berretti, S., Werghi, N., Del Bimbo, A., Pala, P, 2013a. Matching 3D face scans using interest points and local histogram descriptors. Comput. Graph. 37, 509–525.
Berretti, S., Werghi, N., Del Bimbo, A., Pala, P, 2014. Selecting stable keypoints and local descriptors for person identification using 3D face scans. Vis. Comput. 30, 1275–1292.
Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. 14, 239–256.
Bhattacharjee, S., Mohammadi, A., Marcel, S., 2018. Spoofing deep face recognition with custom silicone masks. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems. BTAS, pp. 1–7.
Blanz, V., Vetter, T., 2003. Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. 25, 1063–1074.
Boehnen, C., Peters, T., Flynn, P.J., 2009. 3D signatures for fast 3D face recognition. In: Advances in Biometrics, Third International Conference. ICB 2009.
Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S., 2018. Large scale 3D morphable models. Int J. Comput. Vis. 126, 233–254.
Bronstein, A.M., Bronstein, M.M., Kimmel, R., 2005. Three-dimensional face recognition. Int J. Comput. Vis. 64, 5–30.
Cai, Y., Lei, Y., Yang, M., You, Z., Shan, S., 2019. A fast and robust 3D face recognition approach based on deeply learned face representation. Neurocomputing 363, 375–397.
Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K., 2014. Facewarehouse: A 3D facial expression database for visual computing. IEEE Trans. Vis. Comput. Gr. 20, 413–425, [dataset].
Chang, K.I., Bowyer, K.W., Flynn, P.J., 2005a. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'05 - Workshops, p. 157.
Chang, K., Bowyer, K., Flynn, P., 2005b. Effects on facial expression in 3D face recognition. In: Biometric Technology for Human Identification II. pp. 132–143.
Chang, K.I., Bowyer, K.W., Flynn, P.J., 2005c. An evaluation of multimodal 2D+3D face biometrics. IEEE Trans. Pattern Anal. 27, 619–624, [dataset].
Chang, K.I., Bowyer, K.W., Flynn, P.J., 2006. Multiple nose region matching for 3D face recognition under varying facial expression. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1695–1700.
Chang, F., Tran, A.T., Hassner, T., Masi, I., Nevatia, R., Medioni, G., 2018. ExpNet: Landmark-free, deep, 3D facial expressions. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 122–129.
Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J., 2017. Pointnet: Deep learning on point sets for 3D classification and segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 77–85.
Chen, S., Li, W., Yang, H., Huang, D., Wang, Y., 2020. 3D face mask anti-spoofing via deep fusion of dynamic texture and shape clues. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition. FG 2020, pp. 314–321.
Chen, X., Wu, H., Jin, X., Zhao, Q., 2013. Face illumination manipulation using a single reference image by adaptive layer decomposition. IEEE Trans. Image Process 22, 4249–4259.
Cheng, S., Marras, I., Zafeiriou, S., Pantic, M., 2015. Active nonrigid ICP algorithm. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. FG, pp. 1–8.
Chua, C., Han, F., Ho, Y., 2000. 3D human face recognition using point signature. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). pp. 233–238.
Chung, T.Q., Huyen, H.C., Sang, D.V., 2020. A novel generative model to synthesize face images for pose-invariant face recognition. In: 2020 International Conference on Multimedia Analysis and Pattern Recognition. MAPR, pp. 1–6.
Colbry, D., Stockman, G., 2007. Canonical face depth map: A robust 3D representation for face verification. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7.
Cook, J., Chandran, V., Sridharan, S., 2007a. Multiscale representation for 3-D face recognition. IEEE Trans. Inf. Foren. Sec. 2, 529–536.
Cook, J., Cox, M., Chandran, V., Sridharan, S., 2007b. Robust 3D face recognition from expression categorisation. In: Advances in Biometrics. Berlin, Heidelberg, pp. 271–280.
Corneanu, C.A., Simon, M.O., Cohn, J.F., Guerrero, S.E., 2016. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans. Pattern Anal. 38, 1548–1568.
Dagnes, N., Vezzetti, E., Marcolin, F., Tornincasa, S., 2018. Occlusion detection and restoration techniques for 3D face recognition: a literature review. Mach. Vis. Appl. 29, 789–813.
Danelakis, A., Theoharis, T., Pratikakis, I., 2015. A survey on facial expression recognition in 3D video sequences. Multimed. Tools Appl. 74, 5577–5615.
Deng, J., Cheng, S., Xue, N., Zhou, Y., Zafeiriou, S., 2018. UV-GAN: Adversarial facial UV map completion for pose-invariant face recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7093–7102.
Dou, P., Kakadiaris, I.A., 2018. Multi-view 3D face reconstruction with deep recurrent neural networks. Image Vis. Comput. 80, 80–91.
Dou, P., Shah, S.K., Kakadiaris, I.A., 2017. End-to-end 3D face reconstruction with deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1503–1512.
Dou, P., Wu, Y., Shah, S., Kakadiaris, I., 2014. Robust 3D face shape reconstruction from single images via two-fold coupled structure learning. In: Proceedings of the British Machine Vision Conference.
Drira, H., Amor, B., Daoudi, M., Srivastava, A., 2010. Pose and expression-invariant 3D face recognition using elastic radial curves. In: British Machine Vision Conference.
Ekenel, H.K., Gao, H., Stiefelhagen, R., 2007. 3-D face recognition using local appearance-based models. IEEE Trans. Inf. Foren. Sec. 2, 630–636.
Elaiwat, S., Bennamoun, M., Boussaid, F., El-Sallam, A., 2014. 3-D face recognition using curvelet local features. IEEE Signal Proc. Lett. 21, 172–175.
Elaiwat, S., Boussaid, F., Bennamoun, M., El-Sallam, A., 2013. 3D face identification using curvelet transform. In: 2013 1st International Conference on Communications, Signal Processing, and their Applications. ICCSPA, pp. 1–6.
Emambakhsh, M., Evans, A., 2017. Nasal patches and curves for expression-robust 3D face recognition. IEEE Trans. Pattern Anal. 39, 995–1007.
Emambakhsh, M., Evans, A.N., Smith, M., 2013. Using nasal curves matching for expression robust 3D nose recognition. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems. BTAS, pp. 1–8.
Faltemier, T.C., Bowyer, K.W., Flynn, P.J., 2008. Rotated profile signatures for robust 3D feature detection. In: 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. pp. 1–7.
Feng, Z., Hu, G., Kittler, J., Christmas, W., Wu, X., 2015. Cascaded collaborative regression for robust facial landmark detection trained using a mixture of synthetic and real images with dynamic weighting. IEEE Trans. Image Process. 24, 3425–3440.
Feng, Y., Wu, F., Shao, X., Wang, Y., Zhou, X., 2018. Joint 3D face reconstruction and dense alignment with position map regression network. In: Computer Vision ECCV 2018. Cham, pp. 557–574.
Ganguly, S., Bhattacharjee, D., Nasipuri, M., 2015. Depth based occlusion detection and localization from 3D face image. Int. J. Image Graph. Signal Process. 7, 20–31.
Gawali, A., Deshmukh, R., 2014. 3D face recognition using geodesic facial curves to handle expression, occlusion and pose variations. Int. J. Comput. Sci. Inf. Technol. 5, 4284–4287.

Gilani, S.Z., Mian, A., 2016. Towards large-scale 3D face recognition. In: 2016 International Conference on Digital Image Computing: Techniques and Applications. DICTA, pp. 1–8.
Gilani, S.Z., Mian, A., 2018. Learning from millions of 3D scans for large-scale 3D face recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1896–1905.
Gilani, S.Z., Mian, A., Eastwood, P., 2017. Deep, dense and accurate 3d face correspondence for generating population specific deformable models. Pattern Recognit. 69, 238–250.
Goswami, G., Vatsa, M., Singh, R., 2014. Rgb-d face recognition with texture and attribute features. IEEE Trans. Inf. Foren. Sec. 9, 1629–1640, [dataset].
Guha, R., 2021. A report on automatic face recognition: Traditional to modern deep learning techniques. In: 2021 6th International Conference for Convergence in Technology. I2CT, pp. 1–6.
Guo, B., Da, F., 2019. Expression-invariant 3D face recognition based on local descriptors. J. Comput.-Aided Des. Comput. Graph. 31, 1086–1094.
Guo, M., Da, F., Deng, X., Gai, S., 2017. 3D face recognition based on keypoints and local feature. J. Zhejiang Univer.(Engineering Science) 51, 584–589.
Guo, G., Zhang, N., 2019. A survey on deep learning based face recognition. Comput. Vis. Image Und. 189, 102805.
Guo, Z., Zhang, Y., Xia, Y., Lin, Z., Fan, Y., Feng, D.D., 2013. Multi-pose 3D face recognition based on 2D sparse representation. J. Vis. Commun. Image R. 24, 117–126.
Gupta, S., Aggarwal, J.K., Markey, M.K., Bovik, A.C., 2007. 3D face recognition founded on the structural diversity of human faces. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7.
Han, H., Shan, S., Chen, X., Gao, W., 2008. Illumination transfer using homomorphic wavelet filtering and its application to light-insensitive face recognition. In: 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. pp. 1–6.
Hariri, W., Tabia, H., Farah, N., Benouareth, A., Declercq, D., 2016a. 3D face recognition using covariance based descriptors. Pattern Recognit. Lett. 78, 1–7.
Hariri, W., Tabia, H., Farah, N., Declercq, D., Benouareth, A., 2016b. Hierarchical covariance description for 3D face matching and recognition under expression variation. In: 2016 International Conference on 3D Imaging. IC3D, pp. 1–7.
Hassner, T., Harel, S., Paz, E., Enbar, R., 2015. Effective face frontalization in unconstrained images. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 429, pp. 5–4304.
Hsu, G.J., Liu, Y., Peng, H., Wu, P., 2014. Rgb-d-based face reconstruction and recognition. IEEE Trans. Inf. Foren. Sec. 9, 2110–2118.
Hsu, G., Shie, H., Hsieh, C., Chan, J., 2018. Fast landmark localization with 3D component reconstruction and CNN for cross-pose recognition. IEEE Trans. Circ. Syst. Vid. 28, 3194–3207.
Hu, G., Mortazavian, P., Kittler, J., Christmas, W., 2013. A facial symmetry prior for improved illumination fitting of 3D morphable model. In: 2013 International Conference on Biometrics. ICB, pp. 1–6.
Huang, D., Ardabilian, M., Wang, Y., Chen, L., 2012. 3-D face recognition using eLBP-based facial description and local feature hybrid matching. IEEE Trans. Inf. Foren. Sec. 7, 1551–1565.
Huang, Y., Wang, Y., Tan, T., 2006. Combining statistics of geometrical and correlative features for 3D face recognition. In: Proceedings of the British Machine Vision Conference 2006.
Inan, T., Halici, U., 2012. 3-d face recognition with local shape descriptors. IEEE Trans. Inf. Foren. Sec. 7, 577–587.
Jackson, A.S., Bulat, A., Argyriou, V., Tzimiropoulos, G., 2017. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 1031–1039.
Jahanbin, S., Choi, H., Bovik, A.C., Castleman, K.R., 2007. Three dimensional face recognition using wavelet decomposition of range images. In: 2007 IEEE International Conference on Image Processing. pp. 145–148.
Jan, A., Ding, H., Meng, H., Chen, L., Li, H., 2018. Accurate facial parts localization and deep learning for 3D facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 466–472.
Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., Gao, W., 2005. Efficient 3D reconstruction for face recognition. Pattern Recognit. 38, 787–798.
Jiang, L., Zhang, J., Deng, B., 2020. Robust RGB-D face recognition using attribute-aware loss. IEEE Trans. Pattern Anal. 42, 2552–2566.
Jin, X., Li, Z., Ning, N., Lu, H., Li, X., Zhang, X., Zhu, X., Fang, X., 2020. Face illumination transfer and swapping via dense landmark and semantic parsing. IEEE Sens. J. 1.
Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y., Karampatziakis, N., Theoharis, T., 2007. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE Trans. Pattern Anal. 29, 640–649.
Kan, M., Shan, S., Chang, H., Chen, X., 2014. Stacked progressive auto-encoders (SPAE) for face recognition across poses. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1883–1890.
Kim, D., Choi, J., Leksut, J.T., Medioni, G., 2016. Accurate 3D face modeling and recognition from RGB-d stream in the presence of large pose changes. In: 2016 IEEE International Conference on Image Processing. ICIP, pp. 3011–3015.
Kim, D., Hernandez, M., Choi, J., Medioni, G., 2017. Deep 3D face identification. In: 2017 IEEE International Joint Conference on Biometrics. IJCB, pp. 133–142.
Kittler, J., Koppen, P., Kopp, P., Huber, P., Rätsch, M., 2018. Conformal mapping of a 3D face representation onto a 2D image for CNN based face recognition. In: 2018 International Conference on Biometrics. ICB, pp. 124–131.
Kollreider, K., Fronthaler, H., Faraj, M.I., Bigun, J., 2007. Real-time face detection and motion analysis with application in liveness assessment. IEEE Trans. Inf. Foren. Sec. 2, 548–558.
Križaj, J., Štruc, V., Dobrišek, S., 2013. Combining 3D face representations using region covariance descriptors and statistical models. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. FG, pp. 1–7.
Kusuma, G.P., Chua, C., 2011. PCA-based image recombination for multimodal 2D+3D face recognition. Image Vis. Comput. 29, 306–316.
Le, H.A., Kakadiaris, I.A., 2019. Illumination-invariant face recognition with deep relit face images. In: 2019 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 2146–2155.
Lee, Y., Chen, J., Tseng, C., Lai, S., 2016. Accurate and robust face recognition from RGB-d images with a deep learning approach. In: British Machine Vision Conference 2016.
Lee, D., Krim, H., 2017. 3D face recognition in the Fourier domain using deformed circular curves. Multidim. Syst. Sign P 28, 105–127.
Lei, Y., Bennamoun, M., Hayat, M., Guo, Y., 2014. An efficient 3D face recognition approach using local geometrical signatures. Pattern Recognit. 47, 509–524.
Lei, Y., Guo, Y., Hayat, M., Bennamoun, M., Zhou, X., 2016. A two-phase weighted collaborative representation for 3D partial face recognition with single sample. Pattern Recognit. 52, 218–237.
Li, X., Da, F., 2012. Efficient 3D face recognition handling facial expression and hair occlusion. Image Vis. Comput. 30, 668–679.
Li, S., Deng, W., 2020. Deep facial expression recognition: A survey. IEEE Trans. Affect. Comput. 1.
Li, H., Huang, D., Lemaire, P., Morvan, J., Chen, L., 2011. Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities. In: 2011 18th IEEE International Conference on Image Processing. pp. 3053–3056.
Li, H., Huang, D., Morvan, J., Chen, L., Wang, Y., 2014. Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. Neurocomputing 133, 179–193.
Li, H., Huang, D., Morvan, J., Wang, Y., Chen, L., 2015. Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors. Int J. Comput. Vis. 113, 128–142.
Li, X., Jia, T., Zhang, H., 2009. Expression-insensitive 3D face recognition using sparse representation. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2575–2582.
Li, X., Komulainen, J., Zhao, G., Yuen, P., Pietikäinen, M., 2016a. Generalized face anti-spoofing by detecting pulse from face videos. In: 2016 23rd International Conference on Pattern Recognition. (ICPR), pp. 4244–4249.
Li, H., Sun, J., Chen, L., 2017a. Location-sensitive sparse representation of deep normal patterns for expression-robust 3D face recognition. In: Conference Proceedings. pp. 234–242.
Li, H., Sun, J., Xu, Z., Chen, L., 2017b. Multimodal 2D+3D facial expression recognition with deep fusion convolutional neural network. IEEE Trans. Multimedia 19, 2816–2831.
Li, Y., Wang, Y., Liu, J., Hao, W., 2018. Expression-insensitive 3Dface recognition by the fusion of multiple subject-specificcurves. Neurocomputing 275, 1295–1307.
Li, L., Xu, C., Tang, W., Zhong, C., 2008. 3D face recognition by constructing deformation invariant image. Pattern Recognit. Lett. 29, 1596–1602.
Li, B.Y.L., Xue, M., Mian, A., Liu, W., Krishna, A., 2016b. Robust RGB-D face recognition using kinect sensor. Neurocomputing 214, 93–108.
Li, X., Zhang, H., 2007. Adapting geometric attributes for expression-invariant 3D face recognition. In: IEEE International Conference on Shape Modeling and Applications 2007. SMI '07, pp. 21–32.
Li, Xiaoli, et al., 2010. A rapid method for 3D face recognition based on rejection algorithm. Acta Autom. Sin. (1), 153–158.
Lin, W., Boston, N., Hu, Y.H., 2005. Summation invariant and its applications to shape recognition. In: Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. pp. 205–208.
Lin, Y., Liang, S., Shengjin, W., 2014. 3-d faces registration via non-rigid ICP. J. Tsinghua Univ.(Science and Technology) 54, 334–340.
Lin, S., Liu, F., Liu, Y., Shen, L., 2019. Local feature tensor based deep learning for 3D face recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2019, pp. 1–5.
Lin, T.H., Shih, W.P., Chen, W.C., Ho, W.Y., 2006a. 3D face authentication by mutual coupled 3D and 2D feature extraction. In: Proceedings of the 44st Annual Southeast Regional Conference. pp. 423–427.
Lin, W., Wong, K., Boston, N., Hu, Y.H., 2006b. Fusion of summation invariants in 3D human face recognition. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'06, pp. 1369–1376.
Liu, R., Hu, R., Yu, H., 2014. Nose detection on 3D face images by depth-based template matching. In: 2014 7th International Congress on Image and Signal Processing. pp. 302–307.

Liu, S., Lan, X., Yuen, P.C., 2021. Multi-channel remote photoplethysmography correspondence feature for 3D mask face presentation attack detection. IEEE Trans. Inf. Foren. Sec. 16, 2683–2696.
Liu, Q., Sun, J., 2015. 3D face recognition based on local radial binary pattern. J. Beijing Univer. Aeronaut. Astronaut. 41, 732–736.
Liu, P., Wang, Y., Huang, D., Zhang, Z., 2012. Recognizing occluded 3D faces using an efficient ICP variant. In: 2012 IEEE International Conference on Multimedia and Expo. pp. 350–355.
Liu, P., Wang, Y., Huang, D., Zhang, Z., Chen, L., 2013. Learning the spherical harmonic features for 3-D face recognition. IEEE Trans. Image Process. 22, 914–925.
Liu, S., Yuen, P.C., Zhang, S., Zhao, G., 2016. 3D mask face anti-spoofing with remote photoplethysmography. In: Computer Vision. ECCV 2006, Cham, pp. 85–100.
Liu, F., Zeng, D., Li, J., Zhao, Q., 2017. On 3D face reconstruction via cascaded regression in shape space. Front. Inform. Tech. El 18, 1978–1990.
Liu, Y., Zeng, J., Shan, S., Zheng, Z., 2018a. Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 458–465.
Liu, F., Zhao, Q., Liu, X., Zeng, D., 2020a. Joint face alignment and 3D face reconstruction with application to face recognition. IEEE Trans. Pattern Anal. 42, 664–678.
Liu, F., Zhao, Q., Liu, X., Zeng, D., 2020b. Joint face alignment and 3D face reconstruction with application to face recognition. IEEE Trans. Pattern Anal. 42, 664–678.
Liu, F., Zhu, R., Zeng, D., Zhao, Q., Liu, X., 2018b. Disentangling features in 3D face shapes for joint face reconstruction and recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5216–5225.
Lu, X., Colbry, D., Jain, A.K., 2004. Three-dimensional model based face recognition. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, pp. 362–366.
Lu, X., Jain, A.K., 2005. Deformation analysis for 3D face matching. In: 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Vol. 1. pp. 99–104.
Lu, X., Jain, A.K., 2006. Automatic feature extraction for multiview 3D face recognition. In: 7th International Conference on Automatic Face and Gesture Recognition. FGR06, pp. 585–590.
Lu, X., Jain, A., 2008. Deformation modeling for robust 3D face matching. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1346–1357.
Lu, X., Jain, A.K., Colbry, D., 2006. Matching 2.5d face scans to 3D models. IEEE Trans. Pattern Anal. Mach. Intell. 28, 31–43, [dataset].
Luthi, M., Gerig, T., Jud, C., Vetter, T., 2018. Gaussian process morphable models. IEEE Trans. Pattern Anal. 40, 1860–1873.
Lv, S., Da, F., Deng, X., 2015. A 3D face recognition method using region-based extended local binary pattern. In: 2015 IEEE International Conference on Image Processing. ICIP, pp. 3635–3639.
Ma, M., Hu, X., Xu, Y., Peng, S., 2014. A lighting robust fitting approach of 3D morphable model using spherical harmonic illumination. In: 2014 22nd International Conference on Pattern Recognition. pp. 2101–2106.
Maatta, J., Hadid, A., Pietikainen, M., 2011. Face spoofing detection from single images using micro-texture analysis. In: 2011 International Joint Conference on Biometrics. IJCB, pp. 1–7.
Masi, I., Chang, F., Choi, J., Harel, S., Kim, J., Kim, K., Leksut, J., Rawls, S., Wu, Y., Hassner, T., AbdAlmageed, W., Medioni, G., Morency, L., Natarajan, P., Nevatia, R., 2019. Learning pose-aware models for pose-invariant face recognition in the wild. IEEE Trans. Pattern Anal. 41, 379–393.
Meyer, G.P., Do, M.N., 2018. Real-time 3D face verification with a consumer depth camera. In: 2018 15th Conference on Computer and Robot Vision. CRV, pp. 71–79.
Mian, A.S., Bennamoun, M., Owens, R.A., 2005. Matching tensors for pose invariant automatic 3D face recognition. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05 - Workshops. p. 120.
Mian, A., Bennamoun, M., Owens, R., 2007. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. IEEE Trans. Pattern Anal. 29, 1927–1943.
Mian, A.S., Bennamoun, M., Owens, R., 2008. Keypoint detection and local feature matching for textured 3D face recognition. Int J. Comput. Vis. 79 (1), 1–12. http://dx.doi.org/10.1007/s11263-007-0085-5.
Ming, Y., 2014. Rigid-area orthogonal spectral regression for efficient 3D face recognition. Neurocomputing 129, 445–457.
Ming, Y., 2015. Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. Image Vis. Comput. 35, 14–22.
Ming, Y., Ruan, Q., 2012. Robust sparse bounding sphere for 3D face recognition. Image Vis. Comput. 30, 524–534.
Mohammadzade, H., Hatzinakos, D., 2013. Iterative closest normal point for 3D face recognition. IEEE Trans. Pattern Anal. 35, 381–397.
Mpiperis, I., Malassiotis, S., Strintzis, M.G., 2007. 3-D face recognition with the geodesic polar representation. IEEE Trans. Inf. Foren. Sec. 2, 537–547.
Mpiperis, I., Malassiotis, S., Strintzis, M.G., 2008. Bilinear elastically deformable models with application to 3D face and facial expression recognition. In: 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. pp. 1–8.
Mu, G., Huang, D., Hu, G., Sun, J., Wang, Y., 2019. Led3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 5766–5775.
Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. 24, 971–987.
Pan, G., Han, S., Wu, Z., Wang, Y., 2005. 3D face recognition using mapped depth images. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. p. 175.
Pan, G., Wu, Z., 2005. 3D face recognition from range data. Int. J. Image Graph. 5, 573–593.
Papatheodorou, T., Rueckert, D., 2004. Evaluation of automatic 4D face recognition using surface and texture registration. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. pp. 321–326.
Passalis, G., Kakadiaris, I.A., Theoharis, T., Toderici, G., Murtuza, N., 2005. Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. p. 171.
Passalis, G., Perakis, P., Theoharis, T., Kakadiaris, I.A., 2011. Using facial symmetry to handle pose variations in real-world 3D face recognition. IEEE Trans. Pattern Anal. 33, 1938–1951.
Patil, H., Kothari, A., Bhurchandi, K., 2015. 3-d face recognition: features, databases, algorithms and challenges. Artif. Intell. Rev. 44, 393–441.
Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T., 2009. A 3D face model for pose and illumination invariant face recognition. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 296–301.
Peng, X., Bennamoun, M., Mian, A.S., 2011. A training-free nose tip detection method from face range images. Pattern Recognit. 44, 544–558.
Peng, W., Feng, Z., Xu, C., Su, Y., 2017. Parametric T-spline face morphable model for detailed fitting in shape subspace. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 5515–5523.
Perakis, P., Passalis, G., Theoharis, T., Kakadiaris, I.A., 2013. 3D facial landmark detection under large yaw and expression variations. IEEE Trans. Pattern Anal. 35, 1552–1564.
Piotraschke, M., Blanz, V., 2016. Automated 3D face reconstruction from multiple images using quality measures. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3418–3427.
Prabhu, U., Heo, J., Savvides, M., 2011. Unconstrained pose-invariant face recognition using 3D generic elastic models. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1952–1961.
Qi, C.R., Su, H., Nießner, M., Dai, A., Yan, M., Guibas, L.J., 2016. Volumetric and multi-view CNNs for object classification on 3D data. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 5648–5656.
Queirolo, C.C., Silva, L., Bellon, O.R.P., Segundo, M.P., 2010. 3D face recognition using simulated annealing and the surface interpenetration measure. IEEE Trans. Pattern Anal. 32, 206–219.
Rara, H., Elhabian, S., Starr, T., Farag, A., 2010. 3D face recovery from intensities of general and unknown lighting using partial least squares. In: 2010 IEEE International Conference on Image Processing. pp. 4041–4044.
Rara, H.M., Farag, A.A., Davis, T., 2011. Model-based 3D shape recovery from single images of unknown pose and illumination using a small number of feature points. In: 2011 International Joint Conference on Biometrics. IJCB, pp. 1–7.
Romdhani, S., Ho, J., Vetter, T., Kriegman, D.J., 2006. Face recognition using 3-D models: Pose and illumination. Proc. IEEE 94, 1977–1999.
Romero-Huertas, M., Pears, N., 2008. 3D facial landmark localisation by matching simple descriptors. In: 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems. pp. 1–6.
Samir, C., Srivastava, A., Daoudi, M., 2006. Three-dimensional face recognition using shapes of facial curves. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1858–1863.
Samir, C., Srivastava, A., Daoudi, M., Klassen, E., 2009. An intrinsic framework for analysis of facial surfaces. Int J. Comput. Vis. 82, 80–95.
Samir, C., Vandeborre, J.P., Daoudi, M., 2005. Automatic 3D face recognition using topological techniques. In: 2005 IEEE International Conference on Multimedia and Expo. pp. 450–453.
Sánchez-Escobedo, D., Castelán, M., Smith, W.A.P., 2016. Statistical 3D face shape estimation from occluding contours. Comput. Vis. Image. Und. 142, 111–124.
Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L., 2012. Static and dynamic 3D facial expression recognition: A comprehensive survey. Image Vis. Comput. 30, 683–697.
Saxena, A., Chung, S.H., Ng, A.Y., 2007. 3-d depth reconstruction from a single still image. Int J. Comput. Vis. 76, 53–69.
Smeets, D., Keustermans, J., Vandermeulen, D., Suetens, P., 2013. MeshSIFT: Local surface features for 3D face recognition under expression variations and partial data. Comput. Vis. Image Und. 117, 158–169.
Soltanpour, S., Boufama, B., Jonathan Wu, Q.M., 2017. A survey of local feature methods for 3D face recognition. Pattern Recognit. 72, 391–406.
Soltanpour, S., Wu, Q.M.J., 2019. Weighted extreme sparse classifier and local derivative pattern for 3D face recognition. IEEE Trans. Image Process. 28, 3020–3033.
Sun, L., Pan, G., Wu, Z., Lao, S., 2007. Blinking-based live face detection using conditional random fields. In: Advances in Biometrics. AIB, Berlin, Heidelberg, pp. 252–260.
Tabia, H., Laga, H., Picard, D., Gosselin, P., 2014. Covariance descriptors for 3D shape matching and retrieval. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 4185–4192.

Tan, Y., Lin, H., Xiao, Z., Ding, S., Chao, H., 2019. Face recognition from sequential sparse 3D data via deep registration. In: 2019 International Conference on Biometrics. ICB, pp. 1–8.

Tang, L., Gai, S., Da, F., Deng, X., 2016. 3D face recognition method based on the local binary pattern from vertical and horizontal on the mesh. Chin. J. Sci. Instrum. 37, 1413–1420.

Tang, C., Hsu, G.J., Yap, M.H., 2019. Face recognition with disentangled facial representation learning and data augmentation. In: 2019 IEEE International Conference on Image Processing. ICIP, 167, pp. 0–1674.

Tang, H., Yin, B., Sun, Y., Hu, Y., 2013. 3D face recognition using local binary patterns. Signal Process. 93, 2190–2198.

ter Haar, F.B., Veltkamp, R.C., 2010. Expression modeling for expression-invariant face recognition. Comput. Graph. 34, 231–241.

Theoharis, T., Passalis, G., Toderici, G., Kakadiaris, I.A., 2008. Unified 3D face and ear recognition using wavelets on geometry images. Pattern Recognit. 41, 796–804.

Tian, W., Liu, F., Zhao, Q., 2018. Landmark-based 3D face reconstruction from an arbitrary number of unconstrained images. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 774–779.

Tran, A.T., Hassner, T., Masi, I., Medioni, G., 2017. Regressing robust and discriminative 3D morphable models with a very deep neural network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1493–1502.

Tran, L., Liu, X., 2018. Nonlinear 3D face morphable model. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7346–7355.

Trimech, I.H., Maalej, A., Amara, N.E.B., 2020. Point-based deep neural network for 3D facial expression recognition. In: 2020 International Conference on Cyberworlds. CW, pp. 164–171.

Tsalakanidou, F., Malassiotis, S., Strintzis, M.G., 2005. Face localization and authentication using color and depth images. IEEE Trans. Image Process. 14, 152–168.

Wang, M., Deng, W., 2021. Deep face recognition: A survey. Neurocomputing 429, 215–244.

Wang, S.F., Lai, S.H., 2011. Reconstructing 3D face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2115–2121.

Wang, Y., Liu, J., Tang, X., 2010. Robust 3D face recognition by local shape difference boosting. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1858–1870.

Wang, Y., Pan, G., Wu, Z., 2007. 3D face recognition in the presence of expression: A guidance-based constraint deformation approach. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7.

Wang, Y., Pan, G., Wu, Z., Wang, Y., 2006. Exploring facial expression effects in 3D face recognition using partial ICP. In: Comput. Vis. ACCV 2006. Berlin, Heidelberg, pp. 581–590, [dataset].

Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D., 2009. Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Trans. Pattern Anal. Mach. Intell. 31, 1968–1984.

Wei, X., Longo, P., Yin, L., 2007. Automatic facial pose determination of 3D range data for face model and expression identification. In: Advances in Biometrics. Berlin, Heidelberg, pp. 144–153.

Wen, D., Han, H., Jain, A.K., 2015. Face spoof detection with image distortion analysis. IEEE Trans. Inf. Foren. Sec. 10, 746–761.

Werghi, N., Berretti, S., Del Bimbo, A., 2015a. The mesh-LBP: A framework for extracting local binary patterns from discrete manifolds. IEEE Trans. Image Process. 24, 220–235.

Werghi, N., Tortorici, C., Berretti, S., Bimbo, A.D., 2015b. Representing 3D texture on mesh manifolds for retrieval and recognition applications. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2521–2530.

Wu, Z., Deng, W., 2016. One-shot deep neural network for pose and illumination normalization face recognition. In: 2016 IEEE International Conference on Multimedia and Expo. ICME, pp. 1–6.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3D ShapeNets: A deep representation for volumetric shapes. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1912–1920.

Xu, C., Li, S., Tan, T., Quan, L., 2009. Automatic 3D face recognition from depth and intensity gabor features. Pattern Recognit. 42, 1895–1905.

Xu, C., Tan, T., Li, S., Wang, Y., Zhong, C., 2006a. Learning effective intrinsic features to boost 3D-based face recognition. Comput. Vis. ECCV 2006, 416–427.

Xu, C., Tan, T., Wang, Y., Quan, L., 2006b. Combining local features for robust nose location in 3D facial data. Pattern Recognit. Lett. 27, 1487–1494.

Xu, C., Wang, Y., Tan, T., Quan, L., 2004. Automatic 3D face recognition combining global geometric features with local shape variation information. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. pp. 308–313.

Xu, W., Xie, X., Lai, J., 2021. RelightGAN: Instance-level generative adversarial network for face illumination transfer. IEEE Trans. Image Process. 30, 3450–3460.

Yao, C., Wang, S., Zhang, J., He, W., Du, H., Ren, J., Bai, R., Liu, J., 2021. rPPG-based spoofing detection for face mask attack using efficientnet on weighted spatial-temporal representation. In: 2021 IEEE International Conference on Image Processing. ICIP, pp. 3872–3876.

Yu, Y., Da, F., Guo, Y., 2019. Sparse ICP with resampling and denoising for 3D face verification. IEEE Trans. Inf. Foren. Sec. 14, 1917–1927.

Yu, X., Gao, Y., Zhou, J., 2016a. 3D face recognition under partial occlusions using radial strings. In: 2016 IEEE International Conference on Image Processing. ICIP, pp. 3016–3020.

Yu, X., Gao, Y., Zhou, J., 2016b. Boosting radial strings for 3D face recognition with expressions and occlusions. In: 2016 International Conference on Digital Image Computing: Techniques and Applications. DICTA, pp. 1–6.

Zhang, Y., Guo, Z., Xia, Y., Lin, Z., Feng, D.D., 2012. 2D representation of facial surfaces for multi-pose 3D face recognition. Pattern Recognit. Lett. 33, 530–536.

Zhang, H., Han, H., Cui, J., Shan, S., Chen, X., 2018a. Rgb-d face recognition via deep complementary and common feature learning. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. FG 2018, pp. 8–15.

Zhang, H., Li, Q., Sun, Z., 2018b. Joint voxel and coordinate regression for accurate 3D facial landmark localization. In: 2018 24th International Conference on Pattern Recognition. ICPR, pp. 2202–2208.

Zhang, S., Miao, Q., Huang, M., Zhu, X., Chen, Y., Lei, Z., Wang, J., 2019. Pose-weighted gan for photorealistic face frontalization. In: 2019 IEEE International Conference on Image Processing. ICIP, pp. 2384–2388.

Zhang, L., Samaras, D., 2006. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. IEEE Trans. Pattern Anal. Mach. Intell. 28, 351–363.

Zhang, G., Wang, Y., 2011. Robust 3D face recognition based on resolution invariant features. Pattern Recognit. Lett. 32, 1009–1019.

Zhang, F., Zhang, T., Mao, Q., Xu, C., 2018c. Joint pose and expression modeling for facial expression recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3359–3368.

Zhao, M., Mo, R., Zhao, Y., Shi, Z., Zhang, F., 2017. An efficient three-dimensional reconstruction approach for pose-invariant face recognition based on a single view. In: International Conference on Knowledge Science, Engineering and Management. Cham, pp. 422–431.

Zhong, C., Sun, Z., Tan, T., 2007. Robust 3D face recognition using learned visual codebook. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–6.

Zhou, Y., Deng, J., Kotsia, I., Zafeiriou, S., 2019. Dense 3D face decoding over 2500fps: Joint texture & shape convolutional mesh decoders. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1097–1106.

Zhou, S., Xiao, S., 2018. 3D face recognition: a survey. Hum.-Cent. Comput. Inf. 8 (35).

Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z., 2015. High-fidelity pose and expression normalization for face recognition in the wild. CVPR, pp. 787–796.

Zohra, F.T., Rahman, M.W., Gavrilova, M., 2016. Occlusion detection and localization from kinect depth images. In: 2016 International Conference on Cyberworlds. CW, pp. 189–196.

Zou, L., Cheng, S., Xiong, Z., Lu, M., Castleman, K.R., 2007. 3-d face recognition based on warped example faces. IEEE Trans. Inf. Foren. Sec. 2, 513–528.