

REVIEW

Open Access



3D face recognition: a survey

Song Zhou^{1,2} and Sheng Xiao^{1,2*}

*Correspondence:

xiaosheng@hnu.edu.cn

¹ College of Computer
Science and Electronic
Engineering, Hunan
University, South Lushan
Road, Yuelu District,
Changsha 410082, China
Full list of author information
is available at the end of the
article

Abstract

3D face recognition has become a trending research direction in both industry and academia. It inherits advantages from traditional 2D face recognition, such as the natural recognition process and a wide range of applications. Moreover, 3D face recognition systems could accurately recognize human faces even under dim lights and with variant facial positions and expressions, in such conditions 2D face recognition systems would have immense difficulty to operate. This paper summarizes the history and the most recent progresses in 3D face recognition research domain. The frontier research results are introduced in three categories: pose-invariant recognition, expression-invariant recognition, and occlusion-invariant recognition. To promote future research, this paper collects information about publicly available 3D face databases. This paper also lists important open problems.

Keywords: 3D face recognition, 3D face database, Pose-invariant face recognition, Expression-invariant face recognition, Occlusion-invariant face recognition

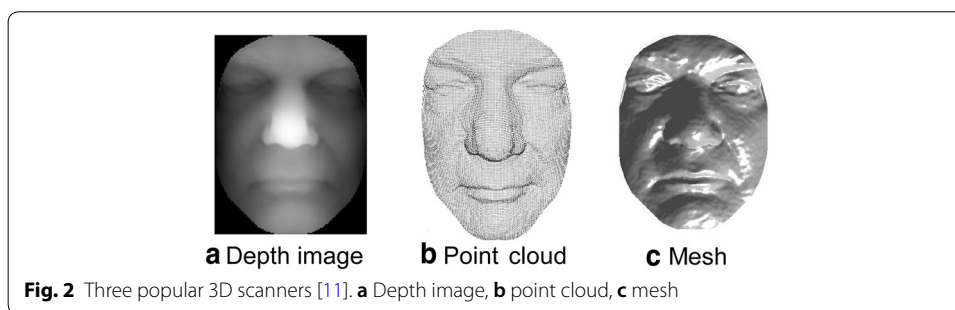
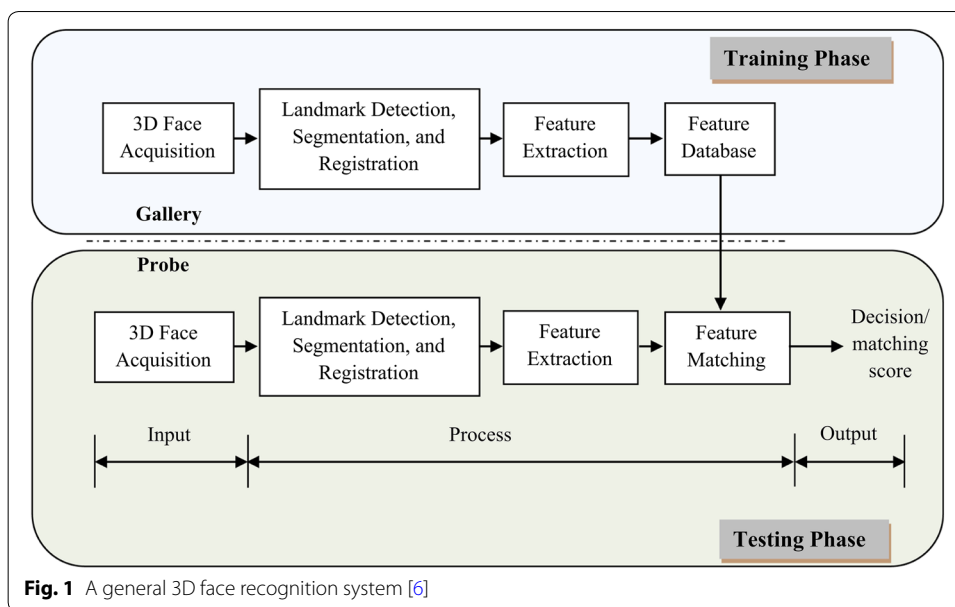
Introduction

Face recognition has been a hot research area for its wide range of applications [1]. In human identification scenarios, facial metrics are more naturally accessible than many other biometrics, such as iris, fingerprint, and palm print [2]. Face recognition is also highly valuable in human computer interaction, access control, video surveillance, and many other applications.

Although 2D face recognition research made significant progresses in recent years, its accuracy is still highly depended on light conditions and human poses [3, 4]. When the light is dim or the face poses are not properly aligned in the camera view, the recognition accuracy will suffer.

The fast evolution of 3D sensors reveals a new path for face recognition that could overcome the fundamental limitations of 2D technologies. The geometric information contained in 3D facial data could substantially improve the recognition accuracy under conditions that are difficult for 2D technologies [5]. Many researchers have turned their focuses to 3D face recognition and made this research area a new trend.

A general work flow for 3D face recognition is shown in Fig. 1. The work flow could be decomposed into two phases and five stages. In the training phase, 3D face data are acquired and then preprocessed to obtain “clean” 3D faces. Then the data are processed by feature extraction algorithms to find the features that could be used to differentiate faces. The features of each face are then stored into the feature database. In the testing phase, the target face goes through the acquisition, preprocessing, and feature extraction



stages that are identical to the stages in the training phase. In the feature matching stage, the features of the target face are compared with the faces stored in the feature database and calculate the match scores. When a match score is sufficiently high, we would claim that the target face is recognized.

3D face acquisition

The acquisition of 3D face samples involves special hardware equipments, which could be categorized as active acquisition systems and passive acquisition systems according to the technologies used. The active acquisition systems actively emit non-visible light, e.g. infrared laser beams, to illuminate the target human face. Then the systems measure the reflection to determine the shape features of the target face. According to the different types of illumination methods, the active acquisition systems could be further categorized as triangulation-based and structured light based. As shown in Fig. 2a, Minolta vivid scanner is an example of triangulation-based 3D scanning system. The scanner measures the emitting and the receiving angles of the laser beam, and then use triangulation methods to determine the exact point of reflection. As the laser beam scanning through the face, a precise map is formed by calculating and grouping many reflection points. The triangulation based systems trade the scanning speed for the precision. It

would require the target man to hold still for several minutes before a 3D face map could be acquired [7]. Therefore this technology is infeasible for the 3D video recording. Compared with the triangulation based systems, the structured light based systems are more popular in consumer level 3D face acquisition. Figure 2b shows a Microsoft Kinect, which emits a light pattern, such as a light grid, to the target face. It then measures the deformation of the light pattern to calculate the surface shape. The structured light based systems offer much faster measurements than the triangulation based systems. However, the structured light measurements often contain holes and artifacts so that the acquired 3D face data are less precise than the triangulation data [8]. Figure 2c shows a Bumblebee XB3, which is a passive acquisition system [9]. It contains several cameras that are placed apart from each other. The system matches points observed from different camera and calculates the exact 3D location of the matched point [10]. The set of the matched points forms the 3D face. Systems like Bumblebee XB3 are often called stereo-imaging systems. Such systems relied on good visible light conditions and usually deliver less precise 3D face data than active 3D face acquisition systems.

Preprocessing

Acquired 3D face data cannot be directly used as the inputs of feature extraction algorithms because the data contain the human faces, but also many distracting features such as hair, ear, neck, eye glasses, and jeweleries. It is true that when us human beings identify each other, these features could be helpful. However, computers are not as intelligent as us at least for now. Features like hair, eye glasses, and jeweleries could be changed from time to time. Ear and neck features are not reliably identifiable for different head poses. These features could be misleading to the current state-of-the-art 3D face recognition algorithms and therefore should be removed before feature extraction.

The first step of preprocessing is to detect the position and orientation of human face. Geometric transformations are used to “turn” the human face to directly against the camera axis. Then the preprocessing uses the help from clearly identifiable facial parts such as nose to isolate the human face area out from areas of the distracting features. This operation is called *segmentation*.

The preprocessed facial data samples are often interpreted in three model formats: depth image, point cloud, and mesh, as shown in Fig. 2. Note that the three model formats are not one to one corresponding to the three popular 3D scanners. They are formats to represent 3D face data.

Feature extraction, feature database, and feature matching

The most straightforward school of feature extraction is to take the entire face as a single feature vector, which is called the global approach [12]. In this approach, the entire face is stored in the database. In the feature matching stage, the target face is compared with faces in database using statistical classification functions [9]. Opposed to the global approach, the component based approach focuses on the local facial characteristics such as nose and eyes. It uses graph operators to extract the nose and eyes part and store these local features in the database. When a target face is inputted for recognition, the component based approach first extract the corresponding parts from the target faces and then searching the matched set of parts from the feature database [13]. There are

hybrid approaches that combine the features used by the global approaches and the local approaches. With more computational cost, the hybrid approach could achieve better recognition accuracy [14].

Methodology

In 3D face recognition system, the selection of feature extraction and matching methods is very important. Both global and local approach have been extensively investigated in the literature and summarized in Table 1.

Performance metrics

This paper proposes the following indications about the performance measures for 3D face recognition tasks.

The notation used for evaluation is as follows:

- TP—the number of samples for the prediction of the positive class as the positive class.
- FN—the number of samples of positive class is predicted to be negative class.
- FP—the number of samples whose negative class is predicted as positive class.
- TN—the number of samples of negative class is predicted to be negative class.

Among them, True and False indicate correct and wrong classification, Positive and Negative samples.

The calculation metrics are as follows:

Accuracy refers to the ratio between the number of samples correctly classified by the classifier and the total number of samples for a given test data set, which reflects the judging ability of the classifier to the entire sample. In other words, it can determine the positive value and the negative value.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

Error rate is the opposite of accuracy rate.

$$Error = \frac{FN + FP}{TP + FN + FP + TN} = 1 - accuracy$$

Precision refers to the proportion of true positive samples in the samples judged as positive by the classifier, that is, how many of all samples judged as positive by the classifier are true positive samples.

Table 1 3D global and local features

3D global features	3D local features
Iterative closest point (ICP)	Landmark-based features
Eigenfaces (PCA)	Curve-based features
Fisherfaces (LDA)	Patch-based features
ICA	

$$Precision = \frac{TP}{TP + FP}$$

Recall refers to the proportion of the positive samples correctly judged by the classifier in the total positive samples, that is, how many of all the positive samples are classified by the classifier as positive samples.

$$Recall = \frac{TP}{TP + FN}$$

Fbeta-score is the harmonic mean of precision and recall.

$$F_{\beta} = (1 + \beta^2) \frac{Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (1)$$

$$= (1 + \beta^2) \frac{(1 + \beta^2)TP}{(1 + \beta^2)TP + \beta^2FP + FN} \quad (2)$$

The value of β ($\beta > 0$) reflects the relative importance of precision and recall in performance evaluation.

When $\beta = 1$, the commonly used F_1 (F measure) value indicates that precision is as important as recall.

$$\frac{2}{F_1} = \frac{1}{Precision} + \frac{1}{Recall}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

$$= \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

The value of F_1 (F measure) is also known as Balanced f-score. When both accuracy and recall are high, the value of F_1 (F measure) is also high.

The rest of this paper is organized as follows: “[History of face recognition research](#)” section introduces significant research results of 3D face recognition in an chronicle order. This help establish a bird-view on this research area. “[Domain research problems](#)” section analyzes current researches and summarize them into domain research problems. “[Research on 3D face databases](#)” section collects the up-to-date information about public 3D face databases, which could facilitate future researches. “[Research on pose-invariant 3D face recognition](#)” section reviews the technologies that could mitigate the pose variation problem for 3D face recognition. “[Research on expression— invariant 3D face recognition](#)” section surveys the technologies that could accurately recognize human faces in different expressions such as laughing or crying, using 3D face information. “[Research on occlusion— invariant 3D face recognition](#)” section reviews the 3D face recognition technologies that could work when the target faces are partially blocked. “[Open problems and perspectives](#)”

section suggest significant problems that are still waiting to be solved in 3D face recognition area. “[Conclusions and discussion](#)” section concludes this paper.

History of face recognition research

Research in face recognition can be dated back to 1960s [15]. From 1964 to 1966 Woodrow W. Bledsoe, along with Helen Chan and Charles Bisson of Panoramic Research, researched programming computers to recognize human faces. Their program asks the administrator to locate the eyes, ears, nose and mouth in the photo. Then, the reference data can be use comparison with the distance and measures. However, because of inconvenience, this work has not received much recognition. Peter Hart at the Stanford Research Institute continued this research, and found optimistic results when using a set of images instead of a set of feature points. Since then, there have been many researches following on this subject and a substantial amount of efforts have been made to find the optimal face recognition method. In the 1970s, Goldstein, Harmon, and Lesk used 21 specific subjective markers such as hair color and lip thickness to automatically identify human faces. The attempt obtained good recognition accuracy. However, the feature measurement and locationing are manually calculated. It is impractical to apply this method to many faces. In 1991, Turk and Pentland proposed a method of using principal component analysis (PCA) to handle face data [16]. This is called the eigenface algorithm which is already become a golden standard for face recognition. Later, inspired by eigenface, a large number of such algorithms were proposed [17–19].

In 1997, Christoph von der Malsburg designed a system that can identify people in photos when the photos are not clear [20]. Followed this work, the research of face recognition diverged into two paths. Face recognition by 3D view is proposed and implemented in systems such as Polar and FaceIt [21].

Although 2D face recognition has achieved considerable success, but the accuracy is still significantly affected by changes in pose and illumination conditions [14, 22]. Many researchers have turned to 3D face recognition because its potential capabilities to overcome the inherent limitations and drawbacks of 2D face recognition. Moreover, the geometric information provided by 3D face data may result in higher recognition accuracy than the 2D case when the pose and illumination conditions are the same [3, 4].

In the late 1980s, [23] used curvature-based methods to test on a small 3D face database, and reached 100% recognition accuracy. In 1996, Gordon’s face recognition experiments showed that combining frontal and side views can improve the recognition accuracy [24]. After that, more and more 3D face recognition research has been proposed, because of the increasing availability of 3D scanning equipments (mainly based on laser and structured light technology).

In 2012, deep learning was first used to analyze and process three-dimensional face images for face recognition [25]. Compared with the traditional method, Deep Convolutional Neural Networks (DCNN) has a great advantage in the processing of image and video, whereas Recurrent Neural Network (RNN) also shows a very good performance in processing continuous data such as voice and text [26]. By using deep learning to train large-scale face datasets, the recognition accuracy of 2d face recognition has been significantly improved [27]. The method of deep learning needs to large datasets to learn face features and be able to depict rich internal information of data. Large-scale 2D face

datasets can be obtained from the Internet. Compared 2D face dataset, training discriminative deep features for 3D face recognition is very difficult due to the lack of large-scale 3D face datasets [27]. In order to solve this problem, Kim et al. [27] proposed using the existing trained 2D face model, and adjust a small amount of 3D face datasets to 3D surface matching. Also, [28] proposed a method for generating a large corpus of labeled 3D face identities and their multiple instances for training and a protocol for merging the most challenging existing 3D datasets for testing. They also proposed the first deep CNN model designed specifically for 3D face recognition and trained on 3.1 million 3D facial scans of 100,000 identities. The proposed training and test datasets are several orders of magnitude larger than previously existing 3D datasets reported in the literature. Based on the 3D datasets, FR3DNet algorithm has been proposed and achieved great accuracy in closed and open world recognition scenarios [28].

In [14], many identification techniques were surveyed. Face recognition can be divided into three categories based on feature extraction methods used in the identification process: global approach, component-based approach and hybrid approach. In the global approach, the entire face is used as a single feature vector for feature classification. The component-based approach mainly analyzes the local facial features such as nose and eyes. The hybrid approach uses both global and local features. The hybrid approach is very effective when the face is frontal and the expression does not change.

Domain research problems

Compared with other popular biometric identification technologies such as fingerprint, iris and retina based recognition, face recognition can identify a person at greater distance. Therefore, it can be applied to various application scenarios such as crowd monitoring and border control. In many of these application scenarios, the 2D face images cannot be accurately recognized due to variations in facial expressions, head pose, occlusion and other factors. Any of these adversary factors could lead to a sharp decrease in recognition efficiency [29].

In 1999, Blanz and Vetter proposed the 3D deformation model (3DMM) synthesis technique and then use this model for 3D face recognition [30]. However, due to the technical limit of the 3D scanning technology at the time, their 3D deformation model was reconstructed from 2D images. It takes a large amount of computation to reconstruct the 3D model. Many researchers agree that 3DMM play an important role in face recognition, but the computational complexity of the reconstruction process hinders its applicability [14, 31–33]. In 2003, Blanz and Vetter proposed to combine 3DMM with 2D image matching technology in order to recognize faces with various head orientations [34]. Unlike [30], their algorithm automatically evaluates all 3D scene parameters, including the position and orientation of the head. Through this new initialization process, the robustness and reliability of the face recognition system is significantly improved. It is noteworthy that the 2D image synthesized 3D facial model is a compromise when fast 3D scanning technology is not available. As soon as people can directly scan 3D face data, models like 3DMM is no longer in active research.

In 2003, Wu et al. [35] proposed 3D face recognition by extracting multiple horizontal profiles from the facial range data. One pitfall of this method is, the recognition accuracy would decrease significantly when the head pose changes. In [1], Zhang compared

the methods and algorithms for 3D face recognition under pose variations, and tests the maximum angle that can be recognized when pose changes. For example, when the face is registered from front and the face model is extracted using the LBP algorithm in [29], an acceptable recognition accuracy could be retained at a maximal face rotation of 60°. Our paper also compares the influence of 2D images and 3D models on recognition performance under changes in head pose. Experiments have shown that 3D models are better tolerant to pose changes than 2D models. We summarized this type of research in “[Research on pose-invariant 3D face recognition](#)” section.

Chua et al. [36] use point signatures in 3D facial recognition. In order to deal with changes in facial expressions, only the rigid part of the face (below the forehead and above the nose) is used. The point signature is also used to locate the reference point in the standardized face model. The images used in the experiment were obtained from the different expressions of 6 subjects, and recognition rate was 100%. The principal component analysis (PCA) method explored by Heshner et al. [37] uses different numbers of feature vectors and image sizes. The image data set used has 37 subjects, each containing 6 different facial expressions. Using multiple images in the gallery improves the recognition accuracy [38]. Moreno et al. [39] segment the 3D face model using Gaussian curvature and then created a feature vector based on the segmented region for the recognition. This method achieved 78% recognition accuracy in a dataset of 420 faces from 60 people with different facial expressions. Our paper summarizes this type of research in “[Research on expression—invariant 3D face recognition](#)” section.

When the face is partially blocked, the recognition accuracy would suffer. In [40, 41], Martinez et al. divided the face model into small areas and proposed a probabilistic approach to match each area locally. The matching results are then combined for the face recognition. Colombo and Cusano [42] propose to recover the blocked part through algorithms and then use the recovered face data in recognition. This method is also useful when people have decorative objects on their face such as scarf, hat, or eye glasses. Our paper summarizes this type of research in “[Research on occlusion—invariant 3D face recognition](#)” section.

In this paper, we will review the latest solutions and the results achieved from the three classes of face recognition research introduced in sections above. Because these researches are all based on some 3D face datasets. In the following sections, we will firstly summarize the current publicly available 3D face database, including the data type of each database, the number of people being collected, the number of scanned images collected, as well as variations in pose, expression, and occlusion.

Research on 3D face databases

There are many large-scale 2D face databases in the world. These databases provide a common platform to evaluate and compare 2D face recognition algorithms. 3D face databases are less common and smaller in scale. Before 2004, there were few publicly available 3D face databases. In recent years, many research institutes have established different kinds of 3D face databases to test and evaluate their own methods for 3D face recognition. Listed below are some of the published 3D databases (see Table 1) that compare different types of data formats, the number of faces, the number of models, and the types of scanning devices. Tables 2, 3 and 4 show the 3D databases constructed

Table 2 Available 3D face databases

Reference/name	Data type	Texture	Number of subject	Number of images	Scanner
ZJU-3DFED	Mesh	Yes	40	360	–
FSU	Mmesh	No	37	222	Minolta Vivid 700
GavabDB	Mesh	No	61	540	Minolta Vi-700 laser range scanner
FRAV3D	Mesh	Yes	105	–	Minolta Vivid 700 red laser light scanner
BU-3DFE	Mesh	Yes	100	2500	Stereo photography, 3DMD digitizer
Beckman	Mesh	Yes	475	–	CyberWare scanner
UoY	Mesh	Yes	350	5000	Stereo vision 3D camera
FRGC v1.0	Range image	Yes	273	943	Minolta Vivid 3D scanner
FRGC v2.0	Range image	Yes	466	4007	Minolta Vivid 3D scanner
UND	Range image	Yes	277	953	Minolta Vivid 900 range scanner
CASIA	Range image	No	123	4623	Minolta Vivid 910 range scanner
ND2006	Range image	Yes	888	13,450	Minolta Vivid 910 range scanner
MSU	Range image	No	90	533	Minolta Vivid 910 range scanner
SHREC08	Range image	No	61	427	–
3D-TEC	Range image	Yes	214	428	Minolta scanner
SHREC11	Range image	No	130	780	Escaan laser scanner
UMB-DB	Range image	Yes	143	1473	Minolta Vivid 900 laser scanner
Texas 3DFRD	Range image	Yes	118	1140	MU-2 stereo imaging system
Bosphorus	Point cloud	Yes	105	4666	The Inspeck Mega Capturor II 3D scanner
Biometrics	Range image	Yes	277	1906	Minolta Vivid 900 range scanner
BJUT-3D	Mesh	Yes	500	–	CyberWare 3030RGB/PS laser scanner
BU-4DFE	3D video	Yes	101	60,600	Di3D (Dimensional Imaging) dynamic system

Table 3 Expression specific 3D face databases [6]

Name	Expressions
FSU	Neutral, smile, scared, angry, squint, frown
GavabDB	Neutral, smile, accentuated laugh, random gesture
FRGC v 2.0	Neutral, surprise, happy, puffy cheeks, anger, frown
BU3D-FE	Neutral, angry, fear, sadness, disgust, happiness, surprise
CASIA	Neutral, smile, eyes closed, anger, laugh, surprise
FRAV3D	Neutral, smile, open mouth, and gesture
ND2006	Neutral, surprise, sadness, disgust, happiness, undetermined
ZJU-3DFED	Neutral, smile, surprise, sad
Bosphorus	Neutral, happy, anger, disgust, fear, sadness, surprise
UoY	Neutral, eyes closed, eyebrows raised, happy, anger
Texas-3D	Neutral, smile/talk with open/closed eyes and/or open/closed mouth
UMB-3D	Neutral, smile, angry, bored
3D-TEC	Neutral, smile

specifically for recognition algorithms that could adapt to the expression variation, the pose variation, and the occlusion variation.

The FRGC [43] database (as shown in Fig. 3c) has tremendous influence on the development of 3D face recognition algorithms. It is widely accepted as a standard reference database to evaluate the performance of 3D face recognition algorithms.

Table 4 Pose specific 3D face databases [6]

Name	Pose variations
GavabDB	Frontal, left profile, right profile, looking up, lookingdown
CASIA	Frontal, tilt left and right from 20° to 30°, up and down from 20° to 30°, left and right from 20° to 30°, left and right from 50° to 60°, left and right from 80° to 90°
FRAV3D	Frontal looking up and down in X-axis direction, 25° Y-axis right turn, 5° Y-axis left turn, small and severe Z-axis right turn
Bosphorus	Frontal, right-downwards, right-upwards, upwards, downwards, slight upwards and slight downwards, or as represented by exact numerical angles + 10°, + 20°, + 30°, + 45°, + 90°, - 45°, - 90°
UoY	Frontal, up, down

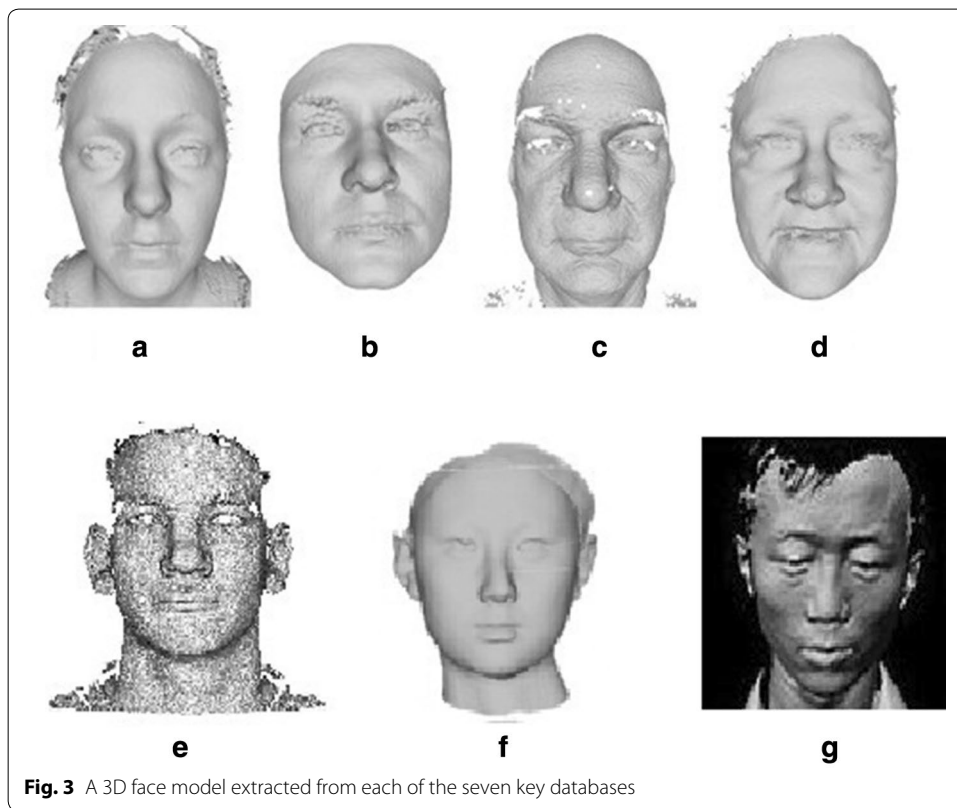


Fig. 3 A 3D face model extracted from each of the seven key databases

The pictures in the database are all 640 * 480 pixel 3D images, scanned by the Minolta Vivid 3D scanner with corresponding RGB texture information. The data was divided into the training set FRGC v1.0, which consisted of 943 scanned images of 273 individuals and the training set (FRGC v2.0), which contains 4007 scanned images of 466 individuals with additional expression tags such as anger, happiness, sadness, surprise and disgust.

BU-3DFE is a 3D face database built specifically for the algorithm development on the expression-invariant face recognition [44] (as shown in Fig. 3a). There are 2500 3D scans from 100 individuals using the stereo photography technique. This database contains 6 types of expressions: anger, happiness, sadness, surprise, disgust, and fear. Each type of expression is further tagged with four different levels.

As shown in Fig. 3b, Bosphorus [45] database contains 3D face images with variations on expressions, head poses, and different types of occlusion. This database is based on 4666 3D scan images of 105 individuals and was scanned using an Inspeck Mega Capturor II 3D scanner.

As shown in Fig. 3e, the ND-2006 dataset [46] was the largest 3D face dataset at the time of publication, and it was also a superset of FRGC v2.0. It contains 13,450 3D scan images with 6 different expression tags (neutral, happy, sad, surprised, disgusted, etc.) and was scanned using a Minolta Vivid 910 range scanner. There were a total of 888 different people had been scanned. Each person had been scanned multiple times. The most scanned person appeared 63 times in the database.

The Texas 3D Face Recognition Database (Texas 3DFRD) [47], shown in Fig. 3d, is a set of 1149 pairs of face texture descriptions and scanned images using the MU-2 stereo imaging system. The database includes 105 adult subjects.

BJUT-3D is a large Chinese Face 3D Face Dataset [44] (as shown in Fig. 3f) which includes 500 Chinese people as the subjects. 250 women and 250 men registered their 3D face data in the database. High-resolution human 3D facial data are scanned using a CyberWare 3030 RGB/PS laser scanner.

As shown in Fig. 3g, the CASIA dataset [48] was tested in 2004 using a non-contact 3D digitizer Minolta Vivid 910 range scanner for 4624 scans of 123 people. The data set not only considers single changes in pose, expression, and lighting, but also changes in expression under the same lighting and pose changes under the same expression.

3D-TEC (3D Twins Expression Challenge (3D-TEC) Data Set) [49], this dataset contains 3D facial scans of 107 pairs of twins, that is 214 people, each with a smile and a neutral expression for a total of 428 scans. Although this data set is ten times smaller than the FRGC v2.0 data set, it is still very representative, because it includes twins with different expressions. This database will help promote the development of 3D face recognition technology.

In contrast to 2D face images, 3D models contains the geometry information and are insensitive to pose and lighting changes [50, 51]. There are two kinds of acquisition techniques for acquiring 3D face models: the active acquisition technologies and the passive acquisition technologies. Examples of the active acquisition technologies include triangulation and structured light. The most typical passive acquisition system is a stereo camera [9]. In active acquisition techniques, such as the Minolta Vivid scanners (shown in Fig. 4a), triangulation technology is used. The scanner emits laser light on the face and then uses the camera to record the image of the light spot. Once the center pixel of the point is calculated, the position of the laser spot is determined by the triangle formed by the laser spot, camera, and laser emitter. The effective range of the triangulation technique could be a few meters with the accuracy of several millimeters. However, the triangulation process could be time-consuming. The scanner has to reconstruct the 3D face model point by point. Using structured light technology, such as the Microsoft Kinect (shown in Fig. 4b), the scanner projects a pattern onto the face surface, and then a camera captures the pattern deformed by the face contour. The shape of the face is calculated based on the deformation of the pattern. Structured light can acquire 3D face data in real time, but the acquired data may contain a large number of holes and artifacts. For a typical passive acquisition system, such as the Bumblebee XB3 (shown in

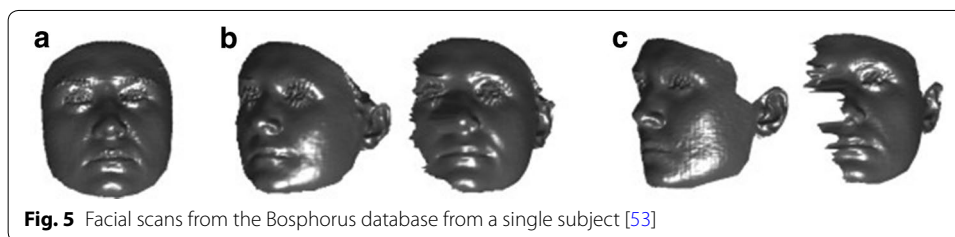
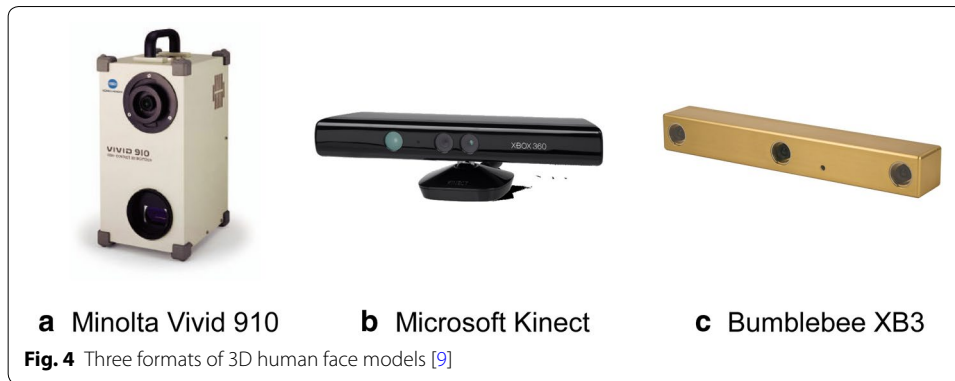


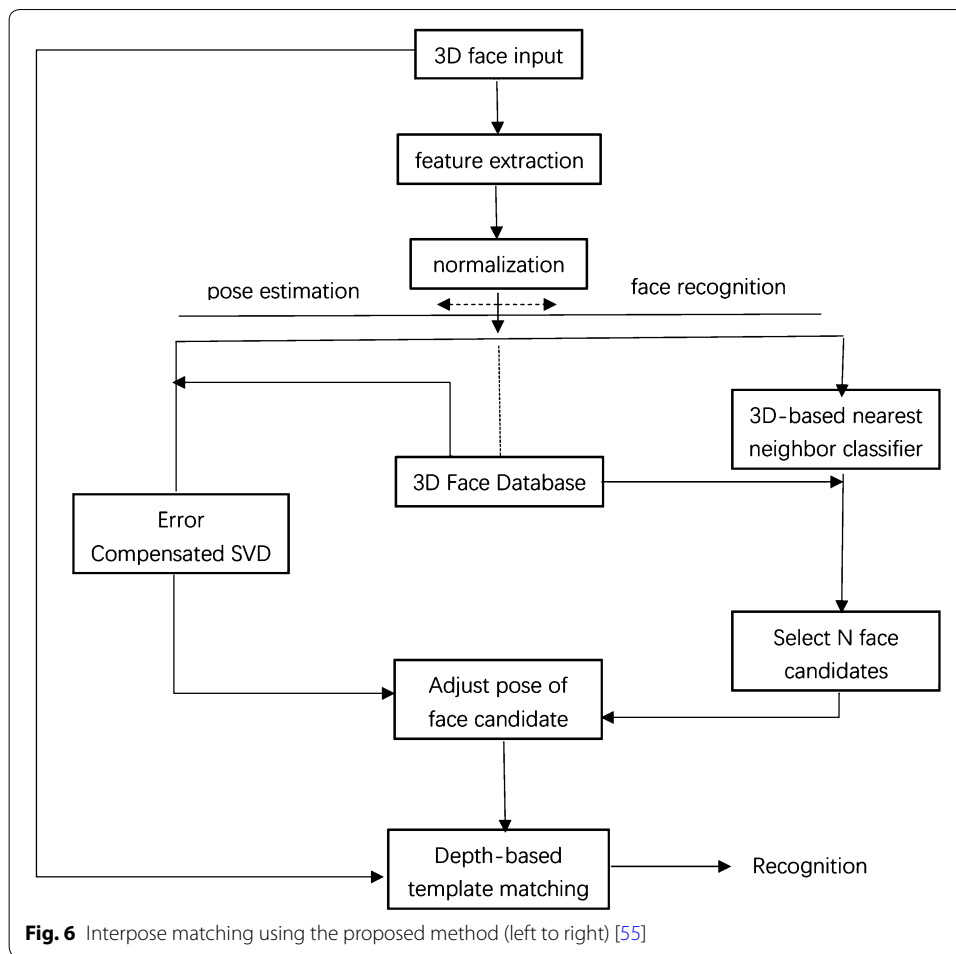
Fig. 4c), the scanner uses two (or more) cameras to take pictures for the face from different angles. The system uses algorithms to match feature points in different pictures and then calculates the exact position of the feature points with the triangulation algorithm. Multiple feature points are calculated simultaneously and then used to reconstruct the 3D face model [10, 52]. The main pitfall of the stereoscopic system is the relatively low resolution of the reconstructed 3D face scans.

3D face recognition algorithms have different performances on different 3D face databases. Many methods are implemented on a specific 3D face database, and performance on other databases may vary.

Research on pose-invariant 3D face recognition

As shown in Fig. 5, in 3D face recognition, the change of head poses can substantially affect the accuracy of 3D face recognition. Many 3D face recognition systems rely on the front face model. Once the head is not upright or the face orientation is rotated away from the front-facing pose, the system would have difficulty to match the face scan with the preset face models.

As early as 2003, Song et al. [54] proposed a 3D face recognition method which could stand with large head deflection. The method depends on the geometric information of the feature points on the face to “adjust” the head pose in the scanned image. Figure 6 briefly shows the extraction of facial feature points, the determination of the head position, and the process of recognition. First, the maximum and minimum curvature points are automatically extracted using the geometric information of the face. These points are composed of the bump points and the nasal peak point (NPP). In order to find the exact position of the head and the deflection angle of the head from the input 3D head image, They proposed the Error Compensated SVD (EC-SVD) algorithm to minimize the least



square error and then compensate in the established 3D normalized space. For each axis, the pose is optimized from the angle acquired by the SVD method, thereby restoring the face model to the frontal angle.

Passalis et al. [55] proposed a method to use face symmetry to resolve the pose variation problem. This method uses wavelet biometric signatures which is also used in the landmark detection algorithms proposed in [56]. The signatures allows a matching for the face symmetry to compensate the pose variation (as shown in Fig. 7). Experiments show that this method is suitable for practical scenarios because it requires no manual intervention and the whole process is fully automatic. Moreover, this method is good at handling extreme pose changes such as a nearly 90° head rotation and leaving only one side of the face to the front.

Perakis et al. [56] proposed an algorithm to handle internal occlusion. The algorithm is based on the annotated face model (AFM). The geometry created by the AFM is also invariant in the event of data missing. Therefore, this method deals with incomplete data problems due to pose changes. Verification experiments had been conducted on FRGC v2.0 and can UND. The UND45LR contains a set of scans with each person turns its head 45° away from the frontal orientation. For each person in the scan, the left pose scan belongs to the training set and the right pose scan is considered to be in the testing set. Similarly, the UND60LR marks a collection of side scans with a 60° pose.

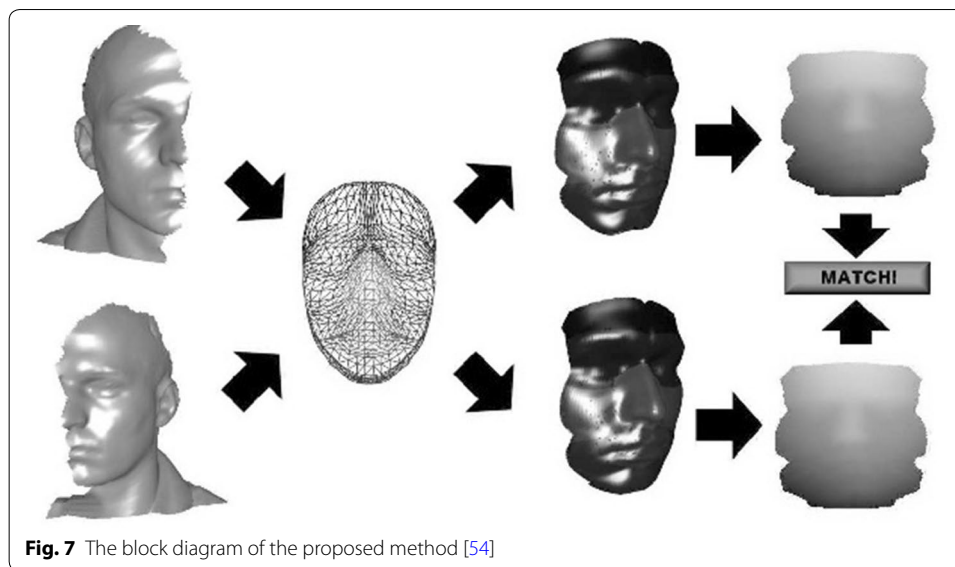


Fig. 7 The block diagram of the proposed method [54]

In [13], a new 3D surface representation method, namely the multi-scale local binary model (MS-LBP) depth map, is proposed. This method is used in conjunction with the shape index (SI) map to increase the significance of the smooth-range surface. Scale Invariant Feature Transform (SIFT) are introduced to extract local features to enhance their robustness to pose variations. The Rank-one recognition rate achieved on the FRGC v2.0 database is 96.1%. Since local facial features are used, this method has been shown to be capable of handle partially occluded facial probes.

Berretti et al. [57] uses Scale Invariant Feature Transform (SIFT) key point detection methods to locate feature points in the depth image and find facial curves that connect these key points. The authors use 45° and 60° side scans in the UND database to test their proposal. Since the same organization has collected UND and FRGC v2.0 databases, they have found 39 identical faces between UND's 45° lateral face and the frontal scan of FRGC v2.0. In addition, there are 33 identical faces in the 60° side scan of the UND, and the frontal face of the FRGC v 2.0 uses the curvature information of the landmark to achieve matching.

In [58], face models are represented by radial curves. In order to overcome the data missing problem caused by pose variation, they used a statistical model in the radial curve's shape space. This method works well for recognition and can reach 98.36% recognition accuracy for faces looking downwards at 35°. However, the scanning result from the right side of the face shows that the recognition rate of the right side scan has dropped to 70.49%, while the left side scan has 86.89%. In addition, the limitation of this method is that manual annotation of the nose tip of the side scan is required.

Mahmood et al. [59] proposed a matching method using nose region extraction to defend against large yaw changes (approximately 60° of yaw axis). In order to re-align the face to the frontal orientation, a pre-defined and pre-trained nose model is used. Face surfaces are represented by local shape descriptors. The effectiveness of this method has been evaluated in the GAVADB 3D facial database, which includes

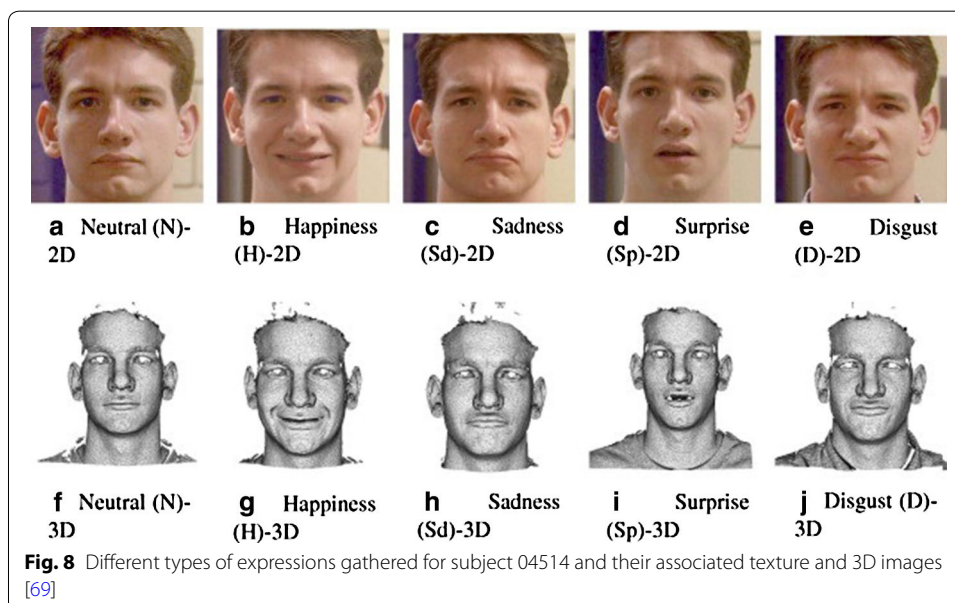
both frontal and partially frontal facial scans. Using this method, the recognition accuracies for frontal face scans and partially frontal facial scans are 94% and 90% respectively.

Ding et al. [60] proposed a PBPR face representation scheme based on the unobstructed facial texture. PBPR can be applied to face images of arbitrary poses, which has greater advantages than other methods. At the same time, they proposed the MtFTL model for learning compact feature transformation between poses.

Research on expression—**invariant 3D face recognition**

Human faces have local non-rigid deformation when the expression changes, which reduces the similarity between the scanned face and the trained face models, and thereby reducing the accuracy of the 3D face recognition algorithms [68]. Figure 8 shows the facial shapes of the five typical expressions: the neutral expression, happiness, sadness, surprise and disgust in 2D and 3D.

3D face recognition methods that can handle the expression changes generally fall into two categories: rigid [4, 69, 70] and non-rigid [71–73]. The rigid method treats the human face as a rigid subject. Such methods are popular in the early days. The main idea is: when the facial expression changes, there are always some facial regions remain unchanged or have little change. These regions are considered as the rigid areas. The features of the rigid areas are extracted and used in face recognition [74]. The most commonly used rigid areas are the nose, eyes, and the area near the forehead. Queirolo and Silva [75] uses the round area around the nose, the ellipse area around the nose, the face area above the nose, and the entire face area to match. The comprehensive four-part score is used to calculate the similarity between two 3D face images. An modified Simulate Annealing Algorithm (SAA) is then used to find the optimal value of the score. This method was tested with the FRGC v2.0 database and achieved a recognition accuracy of 98.4%. Bornak and Rafiei [76] uses the nose area for 3D face recognition. The authors



proposed to firstly search for the nasal area in the center of the image, and then extract the outline of the diagonal area of the nasal area as a feature. Erdogmus et al. [77] proposed another local feature based method. They divided the face into several parts, and then calculated the similarity of the corresponding parts between two 3D face images. The conditional density is used to transform the face recognition problem into a probability optimization problem. Miao and Krim [78] uses the nearest point alignment and level set method to search for a region where two face images matching with each other and then uses the size of the matching areas as the similarity of two face images. The method based on the rigid area is relatively simple and easy to implement. However, this type of method discards areas affected by expressions and does not use all the information contained in the 3D face data.

The non-rigid method applies the deformation recovery algorithms to the 3D facial scan to counteract the distortion caused by expression variations. Although a good recognition method can be found in both categories, the non-rigid method is more capable of handling 3D face recognition in facial expression variations and can extract richer facial information [73]. In non-rigid classification, the recognition algorithms are divided into two categories: local methods and holistic methods.

Local feature based expression—invariant approaches

To our best knowledge, the first review of a 3D face recognition systems based on local processing was composed by Chang et al. [4].

Samir et al. proposed a method of comparing facial shapes by the surface curvature [79]. The basic idea is to roughly represent a facial surface with a limited level curve. The curve is extracted from the depth image. In [80], there is a description of the metric for the facial curve calculation. Experimental results show that this method is robust to various expressions.

Kin-Chung et al. proposed a 3D face recognition system that combines linear discriminant analysis (LDA) and linear support vector machine (LSVM) [81]. This method can obtain the sum of invariants by capturing local characteristics from multiple regions. Ten sub-regions and subsequent feature vectors are extracted from the frontal face image. In addition, the amount of variation is summed using the moving frame technique [82]. LDA and LSVM based on linear optimal fusion rules provide better performance. The performance of the reporting method decreases with the expression increases.

Faltemier et al. used 28 best-performing facial sub-regions for 3D facial recognition [83]. To detect the image, a total of 38 sub-regions were extracted, some of which were overlapping. By using an ICP algorithm, each sub-regions in the probe image matches a gallery image region. The highest Rank-one recognition rate reached 90.2% through single-region matching, which promoted the use of fusion strategies. The improved Borda-count fusion method yields an overall 97.2% Rank-one recognition rate. Although the facial information of the image is not complete in some areas of the FRGC v2.0 database, this algorithm still performs well.

In [84], they proposed a mesh-based 3D face recognition method and evaluated it on the Bosphorus database. The surface micro-components are extracted at the salient points of the local neighborhoods, which are respectively detected by the maximum and minimum curvatures, and the final matching score is determined by the

two salient points. The experimental results on the Bosphorus dataset highlight the effectiveness of the method and its robustness in facial expression variations.

In [85], the meshSIFT algorithm and its 3D face recognition are proposed. The salient points are detected as extremum points in a scale space, and the convex points are determined according to the surface normals in the local neighborhoods that depend on the scale. Position the embossments and describe them in the feature vector of the connected histogram containing the tilt angle and shape index. Because this descriptor is captured from a local area, expressions are almost always preserved. They allow the use of the number of matching features as a measure of similarity to perform 3D face recognition with invariant expression. Using the left-right symmetry of the face to expand the set of feature descriptors, matching features can be found even without overlapping.

Berretti and Werghi [86] proposes a 3D face recognition method based on mesh-DoG keypoints detector and local GH descriptors, and proposes an original solution to improve the stability of keypoints and select the most effective features from local descriptors. Experiments have been conducted to evaluate the effectiveness of optimization recommendations for stable keypoint detection and feature selection. The recognition accuracy was evaluated on the Bosphorus database and the competition results of existing 3D face recognition solutions based on 3D key points are shown.

Tang et al. proposed a local binary model (LBP) based on a 3D facial segmentation scheme [87]. The face surface is divided into 29 sparse areas and 59 dense areas. They used the nearest neighbor to perform 3d face recognition based on classifiers (Table 5).

This paper [88] is based on a new 3D facial feature recognition system, namely Angular Radial Signature(ARS), which is extracted from the semi-rigid region of the face, and then use the kernel principal component analysis (KPCA). The ARSs extract medium characteristics to improve the discriminating ability. The medium features are then connected into a single feature vector, which is input into a Support Vector Machine (SVM) to perform face recognition. This method deals with facial expression changes in different individuals by using face scans. They did a lot of experiments on FGRGC v2.0 and SHREC 2008 data sets to get excellent recognition performance.

Table 5 Occlusion specific 3D face databases [6]

Name	Expressions
FSU	FSU
GavabDB	FSU
FRGC v 2.0	FSU
BU3D-FE	FSU
CASIA	FSU
FRAV3D	FSU
ND2006	FSU
ZJU-3DFED	FSU
Bosphorus	FSU
UoY	FSU
Texas-3D	FSU
UMB-3D	FSU
3D-TEC	FSU

Regional Bounding SpherE descriptors (RBSR) perform effective feature extraction on 3D facial surfaces [89].

In [90] on the application of biometric recognition in 3D face recognition in real life, a new grid SIFT-like algorithm for registration-free 3D face recognition is proposed under expression variations, occlusion, and pose changes. The principal curvature-based 3D keypoint detection algorithm, which can repeatedly recognize the complementary position in the local curvature on a facial scan.

Different region based approaches reported so far are summarized in Table 6.

Holistic approaches

The following table (Table 7) shows the main holistic approaches for using the deformation model functions in different databases.

The method of isometric deformation model belongs to the overall method. In the isometric deformation model method, [96] used the fast-moving method to calculate the geodesic distance of the face surface, established a geodesic distance matrix (GMD), and then used the singular value decomposition (SVD) method to decompose the GMD to obtain the k largest eigenvalues as the shape of the human face descriptor. Miao [97] calculated a set of equal geodesic distance curves for a 3D face surface, and then calculated the evolution vectors between the adjacent two geodesic distance curves. Considering that the evolution vector is easily affected by the deformation of Euclidean space and requires precise face alignment, the author also uses the evolution angle function (EAF) to normalize the evolution vector into a one-dimensional equation. In this way, the comparison problem between two 3D faces is converted into a comparison of two EAF curves. Feng and Krim [98] used the 20 isometric geodesic distance curves from the tip of the nose to represent the human face, and then

Table 6 Pose invariant 3D face recognition approaches

Name	Pose variations	3D face database	No. of faces	Accuracy
Lu et al. [61]	Up to 45°	MSU	300	98%
Dibeklioglu et al. [62, 63]	Up to 45°	Bosphorus	3396	79.41%
Blanz et al. [34, 64]	Up to 40°	FRGC	150	92%
Mian et al. [65]	Up to 90°	FRGC	4007	99.74%
Segundo et al. [66]	< 15°	FRGC 2.0	4950	100%
Wei et al. [67]	< 15°	BU-3DFE	2500	98%
Mahmood et al. [59]	Up to 60°	GavabDB	509	90%
Drira et al. [58]	Up + 35° Down – 35° Right + 90°	GavabDB	549	100% for looking up 98.36% for looking down 70.49% from right side 86.89% from left side
Passalis et al. [55]	Up to 80° along the vertical axis	UND	1018	–
Berretti et al. [57]	Up to 60°	UND FRGC v2.0 GavabDB	4007	82.1%
Perakis et al. [56]	Up to 80° along the vertical axis	UND	414	–
Ding et al. [60]	Up to 90°	LFW	–	92.95%

Table 7 Local processing based 3D face recognition approaches [6]

Category/references	Dataset	Local sub-regions	Identification rate (%)	Verification rate (%)	Classifier/feature detection
Chang et al. [4]	ND2006	Interior nose, center face, entire nose	87.10	–	PCA
Kin-Chung et al. [81]	FRGC v 2.0	10 Sub-regions from nose, eyes, cheeks and eyebrows	–	90.45	LDA,LSVM
Faltemier et al. [83]	FRGC v 2.0	38 Sub-regions from nose, eyes, forehead and chin	97.20	93.2	–
Boehnen et al. [91]	FRGC v 2.0	Eight sub-regions around eyes and nose	95.50	–	Nearest neighbor
Queirolo et al. [75]	FRGC v 2.0	Sub-regions from nose and forehead	98.40	96.50	–
Spreeuwens et al. [92]	FRGC v 2.0	30 Overlapping sub-regions from all over the face	99.00	94.60	PCA-LDA
Li and Da [93]	FRGC v 2.0	Six sub-regions: forehead, left mouth, right mouth, nose, left cheek and right cheek	97.80	96.00	PCA
Lei et al. [94]	FRGC v 2.0	Sub-regions from nose, eyes and forehead area	95.60	97.60	Modified LDA, polynomial-linear-RBF SVM
Tang et al. [87]	BU-3DFE FRGC v2.0	Special feature based sparse division (29 blocks) and dense division (59 blocks)	97.70 94.89	98.20 –	Nearest neighbor
Li et al. [95]	BJUT-3D FRGC v2.0 BU-3DFE Bosphorus	Multi-scale multi resolution patches (MSMC-LNP)	95.10 96.30 92.21 95.40	– – – –	W-SRC
Ming [89]	FRGC v2.0	Seven sub-regions: forehead, mouth, left cheek, right cheek, nose, left eye and right eye regions	–	95.03	–

cut the equal geodesic distance curve into arcs of equal length and then mapped to the Euclidean integral invariant space as face Features to achieve face classification and recognition. This method is tested in the FRGC v2.0 database and the recognition accuracy is 95%.

Berretti and Del Bimbo [99] divide the face surface into a series of equidistant geodesic strips, and then establish a direction index table by measuring the spatial displacement between the equidistant geodesic strips. Finally, compare the orientation index table of the 3D face model to complete face recognition. This form of table-based representation greatly reduces the computational complexity, speeds up the search, and is suitable for large-scale face databases. In addition, some scholars use the elastic geodesic between the facial curves to solve the problem of face expression changes and obtain high recognition accuracy [91, 100].

Table 8 Morphable model based approaches [6]

Category/references	Dataset	Persons in dataset	Images in dataset	Identification rate(%)	Classifier/feature detection
Kakadiaris et al. [72]	FRGC v2.0	466	4007	97.0	–
Xiaoguang and Jain [73]	FRGC v2.0	100	877	92.0	PCA
Mpiperis et al. [101]	BU-3DFE	50	1250	86.0	Maximum likelihood
Amberg et al. [102]	UND	953	953	100	-
	GavabDB	61	427	99.70	
Al-Osaimi et al. [103]	FRGC v2.0	466	4007	96.52	PCA
Haar and Veltkamp [104]	UND	277	953	99.0	PCA
	GavabDB	61	427	98.0	
	BU-3DFE	100	2500	100.0	
	FRGC v2.0	466	4007	97.0	

Mpiperis et al. proposed a geodesic polar coordinatization method for face surfaces [21]. In this way, the internal attributes of the face will not change in the case of isometric deformation, so this representation is suitable for 3D face recognition with expression-invariant. Image classification is done using a PCA classifier and information on colors and shapes is obtained. The experimental results show that the overall performance has been significantly improved by using geodesic polar coordinates.

Research on occlusion—invariant 3D face recognition

However, obtaining non-cooperative individuals' face information in an uncontrolled environment may result in certain parts of the face not being captured because hats, sunglasses, eyes or faces may be partially covered by the hair (Figs. 9, 10). The unavailability of this 3D face data is caused by occlusion of external objects. During the scanning process, due to the non-frontal face pose of the detected individual, some parts of the face may not be captured, which results in erroneous data and we call it internal occlusion. Although many researchers are now dealing with the recognition of expression variations, few researchers do the study of the variation of the occlusion. We will give a detailed introduction to the recognition of some researchers in the case of face occlusion in the following content, including the methods they used, the database they used and the recognition effect that was eventually achieved (Table 8).

Colombo et al. [42] proposed a brand-new recovery strategy that can effectively recognize 3D faces even when faces are partially occlude by unforeseen and unrelated objects (such as scarves, hats, glasses, etc.). The occlusion region is detected by considering their influence on the face projection in a suitable face space. Then, the non-occluded region is used to restore the missing information. Any recognition algorithm can be applied to

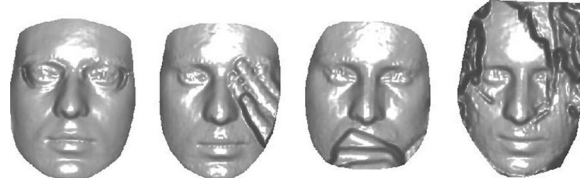


Fig. 9 Four occlusion types in the Bosphorus database [45]



this recovery strategy. This recovery strategy fixes 52 3D faces with all kinds of occlusion and has achieved very good results.

Alyuz et al. [105] proposed a new 3D face registration and recognition method for partial face regions, which can achieve a good recognition effect in the expression and face occlusion. They proposed a fast and flexible alignment method using average regional models (ARMs) to infer local information by iterating the closest point (ICP) algorithm. Different scores from local regional matchers are derived from local regional matchers are fused to robustly identify probe subjects. In this work, a multi-expression 3D facial database and a Bosphorus 3D face database containing a large number of different types of expressions and realistic face occlusion are used for experimental testing. When face were blocked, a good recognition effect was obtained, and the recognition rate increased from 47.05 to 94.12%.

Mayo and Zhang [106] proposed and evaluated a 3D face recognition algorithm based on point cloud rotations, multiple projections, and voted keypoint matching. His basic idea is to rotate every 3D point cloud that represents a person on the x , y , or z -axis, iteratively project 3D points onto multiple 2.5D images in each step of the rotation. The marked keypoint is then extracted from the generated 2.5D image, and this smaller keypoint will replace the original face scan and its projection in the face database. In an extensive assessment using the GavabDB 3D facial recognition data set, their method has a recognition rate of 95% in neutral expressions, and 90% in recognition of faces such as smiles, laughing faces, and partial occlusion of faces.

Alyuz et al. [107] proposed a new type of a novel occlusion-resistant 3D face recognition system that can cope with severe occlusions of hair, hands, and glasses. A two-step registration model first detects the nose region on the curvedness-weighted convex shape index map and then uses the nose-based iterative closest point (ICP) algorithm to perform well alignment. The occlusion region is automatically determined by a generic facial model. After the occluded introduction of the non-facial part is removed, Gappy PCA is used to recover the entire face from the non-occlude facial surface. Experimental results obtained on realistically occluded facial images from the Bosphorus 3D face database show that using the score level fusion of the regional Linear Discriminant Analysis (LDA) classifier, this method improves the Rank-one recognition accuracy significantly from 76.12 to 94.23%.

Alyuz et al. [108] proposed a fully automatic and effective 3D face recognition method, which is robust to face occlusion. In order to align the occluded surfaces, they use a model based registration scheme in which the model is selected to adaptive the face's occlusion. The alignment model is formed by the automatic inspection for validity and

includes only the patch of the non occluded face. By registering the occlusion surfaces of the adaptive selection model, a one-to-one correspondence between the model and non-occlusion surface points is obtained. Therefore, occlusion face registration can be achieved. Compared with the registration strategy based on the overall face model (which is usually used for non-occluded surfaces), the recognition rate of the registration strategy is better than that of the overall face model and achieved about 20% improvement identification rate by the adaptive model method testing of Bosphorus and UMB-DB databases. Drira et al. [58] proposed a new geometric framework for analyzing 3D faces and give a specific targets for comparison, matching, and averaging their shapes. They use radial curves from the tip of the nose to represent facial surfaces and use the elastic shape analysis of these curves to form a Riemannian frame to analyze the shape of the entire facial surface. The representation, together with the elastic Riemannian metric, seems to be naturally used to measure facial deformation and is very robust to partial obstructions and glasses, hair, etc.

Open problems and perspectives

3D face recognition still have a lot of open problems for us to research, such as automatic facial expression recognition, age-invariant face recognition and transfer learning.

As an important part of face recognition technology, facial expression (emotion) recognition (FER) has received extensive attention in the fields of human-computer interaction, security, robot manufacturing, automation, medical care, communication and driving in recent years, and become an active research field in the academic and industrial circles [109]. 3D facial expression recognition can overcome weakness and improve recognition accuracy. Some efforts have focused on the recognition of complex and spontaneous emotions rather than the identification of a typical emotional expression that is deliberately displayed [110–113].

Most face recognition systems are sensitive to age variation. Although some of the earlier papers proposed some recognition methods under age-variations [114–117]. But there are still many problems that need us to explore.

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. The performance of 2D face recognition algorithms has significantly increased by leveraging the representational power of deep neural networks and the use of large-scale labeled training data. As opposed to 2D face recognition, training discriminative deep features for 3D face recognition is very difficult due to the lack of large-scale 3D face datasets. In [27], they show that transfer learning from a CNN trained on 2D face images can effectively work for 3D face recognition by finetuning the CNN with a relatively small number of 3D facial scans.

3D facial expression analysis, recognition under age-variations and transfer learning constitute three open problems that is still in its infancy.

Future 3D technology will be applied to 3D sensing and 3D visualization. 3D sensing is a depth-sensing technology that augments camera capabilities for facial and object recognition in augmented reality, gaming, autonomous driving, and a wide range of applications. 3D sensing technology is about to go full-on mainstream as the likes of Apple, Google and Samsung race to incorporate 3D sensors into their next

generation of smartphones. 3D visualization is the latest mainstream technology that allows designing objects in three-dimensional space with the help of 3D software and producing high-quality digital products. 3D visualization will be used in games, cartoons, films and motion comics because this is the exact sphere for developing and improvement the 3D visualization production.

Conclusions and discussion

3D face recognition is an important and popular area in recent years. More and more researchers are working on this field and presenting their 3D face recognition methods. In this paper, we surveyed some of the latest methods for 3D face recognition under expressions, occlusions, and pose variations. At first we summarized some various available 3D face databases. All of the above methods are tested on these databases. Almost all researchers use the following three formats of face data: point cloud, mesh and range data. All three type face data are obtained by 3D scanner.

The recognition methods are mainly divided into two categories: local methods and holistic methods. Although many experiments are carried out based on the holistic method, we believe that the local method is more suitable for 3D face recognition. Compared to holistic methods, the local method has stronger robustness in terms of occlusion and can obtain better experimental results.

This survey divided 3D face recognition into three directions, pose-invariant 3D face recognition, expression-invariant 3D face recognition and occlusion-invariant 3D face recognition.

This paper survey some methods for pose-invariant face recognition that handles a wide range of poses on publicly available databases. The recognition method is mainly local method. For instance, By using half face matching, a complete face model can be synthesized [55]. Using a statistical model in the radial curve's shape space to overcome the data missing problem.

There are two methods for Expression-Invariant 3d face recognition, one is local approaches based expression invariant approaches and the other is holistic approaches.

This survey made a detailed introduction to the recognition of some researchers in the case of face occlusion in the following content, including the methods they used, the database they used and the recognition effect that was eventually achieved.

Also, 3D face recognition technology has been applied in many fields, such as access control and automatic driving. The iPhone X uses Face ID, technology that unlocks the phone by using infrared and visible light scans to uniquely identify your face. It works in a variety of conditions and is extremely secure. In the world of autonomous driving, the autopilot needs to manage the hand-over between the automated and the manual modes. To have a smooth hand-over, it is important to make sure that the driver is alert and ready to take control of the car before the autopilot is disengaged. To have a smooth transition between modes of operation, Omron introduce 3D facial recognition technology that detects a drowsy or distracted driver. Considering the fact that one out of every six car accidents is attributed to a drowsy or distracted driver, the technology can have a huge impact even on the safety of manual driving.

Expression-invariant and occlusion-invariant 3D face recognition are very active and also proposed a lot of high recognition rates methods. We expect that all three directions can get a well performance in the near future.

Authors' contributions

SX provided guidance for this paper. SZ reviewed a large number of papers, and was a major contributor in writing the manuscript. Both authors read and approved the final manuscript.

Author details

¹ College of Computer Science and Electronic Engineering, Hunan University, South Lushan Road, Yuelu District, Changsha 410082, China. ² National Collaborative Innovation Center for High Performance Computing, Beijing, China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analysed during this study are included in this published article [9, 10, 43–52].

Funding

This work is supported in part by National "2011" Collaborative Innovation Center for High Performance Computing, Hunan University Teaching Reform Funds 2017–2018, University-Corporation Collaborative Student Practice Base Fund 2018–2019.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 June 2018 Accepted: 8 November 2018

Published online: 24 November 2018

References

- Zhang X, Gao Y (2009) Face recognition across pose: a review. *Pattern Recognit* 42:2876–2896
- Chellappa R, Wilson SSC (1995) Human and machine recognition of faces: a survey. *IEEE* 83:705–740
- Xu C, Wang Y, Tan T, Quan L (2004) Depth vs. intensity: which is more important for face recognition? *Int Conf Pattern Recognit* 4:342–345
- Bowyer KW, Chang PFK (2006) A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Comput Vis Image Underst* 101:1–15
- Zhu X, Lei Z, Yan J, Yi D, Li SZ (2015) High-fidelity pose and expression normalization for face recognition in the wild. *Computer vision and pattern recognition*. pp 787–796
- Patil H, Kothari KBA (2015) 3-d face recognition: features, databases, algorithms and challenges. *Artif Intell Rev* 44:393–441
- Zaharescu A, Boyer E, Varanasi K, Horaud R (2009) Surface feature detection and description with applications to mesh matching. In: *Proc. IEEE Conf. on Comput*
- Kittler J, Hilton MH (2005) A survey of 3d imaging, modelling and recognition approaches. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. pp 114–114
- Bennamoun M, Guo Y, Sohel F (2015) Feature selection for 2d and 3d face recognition. *Research Gate*, vol 17
- Guo Y, Zhang MLJWYM J (2014) Benchmark datasets for 3d computer vision. In: *Industrial electronics and applications*. pp 1846–1851
- Soltanpour S, Boufama QJWB (2017) A survey of local feature methods for 3d face recognition. *Pattern Recognit* 72:391–406
- Zhu J, San-Segundo R, Pardo JM (2017) Feature extraction for robust physical activity recognition. *Human-centric Comput Inform Sci* 7(1):16
- Huang D, Zhang G, Ardebilian M, Wang Y, Chen L (2010) 3d face recognition using distinctiveness enhanced facial representations and local feature hybrid matching. In: *Biometrics: theory, applications and systems*
- Zhao W, Chellappa PJPARR (2003) Face recognition: a literature survey. *ACM Comput Surv* 35:399–485
- Bledsoe WW (1966) The model method in facial recognition. *Panoramic Research*
- Pentland MT (1991) Face recognition using eigenfaces. *Computer vision and pattern recognition*. pp 586–591
- Belhumeur PN, Hespanha DJKJP (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Trans Pattern Anal Mach Intell* 19:711–720
- Frey BJ, Colmenarez TSH A (1998) Mixtures of local linear subspaces for face recognition. In: *Computer vision and pattern recognition*
- Moghaddam B, Jebara APT (2000) Bayesian face recognition. *Pattern Recognit* 33:1771–1782
- Wiskott L, Fellous JM, Kruger N, Von Der Malsburg C (1997) Face recognition by elastic bunch graph matching. *PAMI* 17:775–779
- Mpiperis I, Malassiotis MGSS (2007) 3-d face recognition with the geodesic polar representation. *Inform Forensics Secur* 5:537–547

22. Abate AF, Nappi DRGSM (2007) 2d and 3d face recognition: a survey. *Pattern Recognit Lett* 28:1885–1906
23. Cartoux JY, LaPrete JT, Richetin M (1989) Face authentication or recognition by profile extraction from range images. *Interpretation of 3D Scenes*. pp 194–199
24. Gordon G (1995) Face recognition from frontal and profile views. In: *International workshop on automatic face and gesture recognition* pp 47–52
25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp 1106–1114
26. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
27. Kim D, Hernandez JC M, Medioni, G (2017) Deep 3d face identification. [arXiv:1703.10714](https://arxiv.org/abs/1703.10714) <http://arxiv.org/abs/1703.10714>
28. Gilani SZ, Mian A (2017) Learning from millions of 3d scans for large-scale 3d face recognition. *ArXiv e-prints*. [arXiv:1711.05942](https://arxiv.org/abs/1711.05942)
29. Prabhu U, Heo MSJ (2011) Unconstrained poseinvariant face recognition using 3d generic elastic models. *Pattern Anal Mach Intell* 33:1952–1961
30. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3d faces. In: *Computer graphics, annual conference series*. pp 187–194
31. Lee KC, Ho J, Yang MH, Kriegman D (2003) Video-based face recognition using probabilistic appearance manifolds. In: *Computer vision and pattern recognition*
32. Hu Y, Jiang D, Yan S, Zhang L (2004) Automatic 3d reconstruction for face recognition. In: *FG* pp 843–848
33. Arandjelovic O, Shakhnarovich G, Fisher J, Cipolla R, Darrell T (2005) Face recognition with image sets using manifold density divergence. In: *Computer vision and pattern recognition* pp 581–588
34. Vetter VBT (2003) Face recognition based on fitting a 3d morphable model. *Pattern Anal Mach Intell* 25:1063–1074
35. Pan G, Wu YPZ (2003) Automatic 3d face verification from range data. *Acoustics Speech Sign Process* 3:193–196
36. Chua CS, Han F, Ho YK (2000) 3d human face recognition using point signature. In: *Autom. Face Gesture Recog.* pp 233–238
37. Heshner C, Srivastava A, Erlebacher G (2003) A novel technique for face recognition using range imaging. In: *Signal processing and its applications*. pp 201–204
38. Min J, Flynn PJ, Bowyer KW (2003) Using multiple gallery and probe images per person to improve performance of face recognition. In: *Notre Dame computer science and engineering technical report*
39. Moreno AB, Sanchez A, Velez JF, Diaz FJ (2003) Face recognition using 3d surface extracted descriptors. In: *Proceedings of the Irish machine vision and image processing*
40. Martínez AM (2000) Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach. In: *Computer vision and pattern recognition*. pp 712–717
41. Tan X, Chen ZHZFZS (2005) Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. *Vision* 16:875–886
42. Colombo A, Cusano C, Schettini R (2006) Detection and restoration of occlusions for 3d face recognition. In: *Multi-media and Expo*. pp 1541–1544
43. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of face recognition grand challenge. In: *Computer vision and pattern recognition*. pp 947–954
44. Yin L, Wei X, Sun Y, Wang J, Rosato MJ (2006) A 3d facial expression database for facial behavior research. In: *Automatic face and gesture recognition*. pp 211–216
45. Savran A, Alyuz N, Dibeklioglu H, Celiktutan O, Gokberk B, Sankur B, Akarun L (2008) Bosphorus database for 3d face analysis. In: *Biometrics and identity management*
46. Faltemier TC, Bowyer KW, Flynn PJ (2007) Using a multi-instance enrollment representation to improve 3d face recognition. In: *Biometrics: theory, applications, and systems*. pp 1–6
47. Gupta S, Castleman KR, Markey MK, Bovik AC (2010) Texas 3d face recognition database. In: *Southwest Symp. image analysis interpretation*. pp 97–100
48. Xu C, Tan T, Li S, Wang Y, Zhong C (2006) Learning effective intrinsic features to boost 3d-based face recognition. In: *Computer vision*. pp 416–427
49. Vijayan V, Bowyer KW, Flynn PJ, Huang D, Chen L, Hansen M, Ocegueda O, Shah SK, Kakadiaris IA (2011) Twins 3d face recognition challenge. In: *International joint conference on biometrics*
50. Guo Y, Sohel FA, Bennamoun M, Wan J, Lu M (2013) Rops: a local feature descriptor for 3d rigid objects based on rotational projection statistics. In: *Signal processing*
51. Guo Y, Wan J, Lu M, Niu W (2013) A parts-based method for articulated target recognition in laser radar data. *Optik* 124:2727–2733
52. Esteban CH, Schmitt F (2002) Multi-stereo 3d object reconstruction. In: *3D data processing visualization and transmission*. pp 159–166
53. Liang Y, Zhang XXZY (2017) Pose-invariant 3d face recognition using half face. *Sign Process* 57:94
54. Song H, Yang KSU (2004) 3d face recognition under pose varying environments. *Lect Notes Comput Sci* 2908:333–347
55. Passalis G, Perakis P, Theoharis T, Kakadiaris IA (2011) Perakis: Using facial symmetry to handle pose variations in real-world 3d face recognition. *Pattern Anal Mach Intell* 33:1938–1951
56. Perakis P, Passalis G, Theoharis T, Toderici G, Kakadiaris IA (2009) Partial matching of interpose 3d facial data for face recognition. In: *Biometrics: theory, applications, and systems*. pp 439–446
57. Berretti S, Del Bimbo PPA (2013) Sparse matching of salient facial curves for recognition of 3d faces with missing parts. *Forensics Secur* 8:374–389
58. Drira H, Amor BB, Srivastava A, Daoudi M, Slama R (2013) 3d face recognition under expressions, occlusions, and pose variations. *Pattern Anal Mach Intell* 35:2270–2283
59. Mahmood SA, Ghani RF, Kerim AA (2014) 3d face recognition using pose invariant nose region detector. In: *Computer science and electronic engineering conference*

60. Hua WG (2009) Implicit elastic matching with random projections for pose-variant face recognition. In: *Comput. Vis. Pattern Recognit.* pp 1502–1509
61. Lu X, Jain AK (2006) Automatic feature extraction for multiview 3d face recognition. In: FG
62. Dibeklioglu H (2008) Part-based, 3d face recognition under pose and expression variations. Master's thesis. Bogazici University
63. Dibeklioglu H, Salah AA, Akarun L (2008) 3d facial landmarking under expression, pose and occlusion variations. In: *Biometrics theory, applications and systems* pp 1–6
64. Blanz V, Scherbaum K, Seidel HP (2007) Fitting a morphable model to 3d scans of faces. In: *Computer vision.* pp 1–8
65. Mian AS, Bennamoun ROM (2007) An efficient multimodal 2d–3d hybrid approach to automatic face recognition. *Pattern Anal Mach Intell* 29:1584–1601
66. Segundo MP, Queirolo C, Bellon OR, Silva L (2007) Automatic 3d facial segmentation and landmark detection. In: *Image analysis and processing.* pp 431–436
67. Wei X, PL, Yin L (2007) Automatic facial pose determination of 3d dynamic range data for face model and expression identification. In: *IEEE/IAPR 2nd international conference on biometrics*
68. Al-Osaimi F, Bennamoun M, Mian A (2009) An expression deformation approach to non-rigid 3D face recognition. *Int J Comput Vis.* 81(3):302–316
69. Faltemier TC, Bowyer PJFKW (2008) Using multi-instance enrollment to improve performance of 3d face recognition, computer vision and image understanding. *Vis Image Understanding* 112:114–125
70. Mian A, Bennamoun M, Owens R (2006) Automatic 3d face detection, normalization and recognition. In: *3D data processing, visualization and transmission*
71. Bronstein AM, Bronstein MM, Kimmel R (2005) Three-dimensional face recognition. *Int J Comput Vis.* 64(1):5–10
72. Kakadiaris IA, Passalis GTMNMVLYNKTG (2007) Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *Pattern Anal Mach Intell* 29:640–649
73. Lu X, Jain A (2006) Deformation modeling for robust 3d face matching. In: *Computer vision and pattern recognition.* pp 1377–1383
74. Amor BB, Ardabilian M, Chen L (2008) Toward a regionbased 3d face recognition approach. In: *Multimedia and expo, hannover.* pp 101–104
75. Queirolo CC, Silva L, Bellon OR, Segundo MP (2010) 3d face recognition using simulated annealing and the surface interpenetration measure. *Pattern Anal Mach Intell* 32:206–219
76. Bornak B, Rafiei S, Sarikhani A, Babaei A (2010) 3d face recognition by used region-based with facial expression variation. In: *Signal processing systems*
77. Erdogmus N, Daniel L, Dugelay JL (2012) Probabilistic fusion of regional scores in 3d face recognition. *Image processing*
78. Miao S, Krim H (2011) Robustness and expression independence in 3d face recognition. In: *Signal processing systems*
79. Samir C, Srivastava A, Daoudi M (2006) Three-dimensional facerecognition using shapes of facial curves. *Pattern Anal Mach Intell.* 28:1858–1863
80. Klassen E, Srivastava A, Mio M, Joshi SH (2004) Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans Pattern Analy Mach Intell* 26(3):372–383
81. Wong KC, Lin YHHNBXZWY (2007) Optimal linear combination of facial regions for improving identification performance. *Syst Man Cybern B* 37:1138–1148
82. Fels M, Olver PJ (1997) Moving coframes. I. A practical algorithm. *Acta Appl Math* 51:99–136
83. Faltemier TC, Bowyer KW, Flynn PJ (2008) A region ensemble for 3d face recognition. In: *Information forensics and security* 3:62–73
84. Li H, Huang PLJ-MMLCD (2011) Expression robust 3d face recognition via mesh-based histograms of multiple order surface differential quantities. *Image Process (ICIP):*3053–3056
85. 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 Underst* 117:158–69
86. Berretti S, Werghe N, Del Bimbo A, Pala P (2014) Selecting stable keypoints and local descriptors for person identification using 3D face scans. *Vis Comput* 30(11):1275–92
87. Tang H, Yin B, Sun Y, Hu Y (2013) 3D face recognition using local binary patterns. *Sign Process* 93(8):2190–8
88. Lei Y, Bennamoun MHYGM (2014) An efficient 3d face recognition approach using local geometrical signatures. *Pattern Recognit* 47:509–524
89. Ming Y (2015) Robust regional bounding spherical descriptor for 3d face recognition and emotion analysis. *Image Vis Comput* 35:14–22
90. Li H, Huang D, Morvan JM, Wang Y, Chen L (2014) Towards 3d face recognition in the real: a registration-free approach using fine-grained matching of 3d keypoint. *Int J Comput Vis* 113:1–14
91. Boehnen C, Peters T, Flynn PJ (2009) 3d signatures for fast 3d face recognition. In: *IAPR/IEEE Int'l Conf. Biometrics.* pp 12–21
92. Spreuwers L (2011) Fast and accurate 3d face recognition. *Int J Comput Vis* 93(3):389–414
93. Li X, Da F (2012) Efficient 3d face recognition handling facial expression and hair occlusion. *Image Vis Comput* 30:668–679
94. Lei Y, Bennamoun M, El-Sallam AA (2013) An efficient 3d face recognition approach based on the fusion of novel local low-level features. *Pattern Recognit* 46:24–37
95. Li H, Huang J-MMLCYWD (2014) Expression-robust 3d face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. *Neurocomputing* 133:179–193
96. Smeets D, Fabry T, Hermans J, Vandermeulen D, Suetens P (2010) Fusion of an isometricdeformation modeling approach using spectraldecomposition and a region-based approach using icp for expression invariant 3d face recognition. In: *International conference on pattern recognition*

97. Miao S, Krim H (2010) 3d face recognition based on evolution of iso-geodesic distance curves. In: Acoustics, speech, and signal processing. pp 1134–1137
98. Feng S, Krim H, Kogan IA (2007) 3d face recognition using euclidean integral invariants signature. In: Statistical signal processing, pp 156–160
99. Berretti S, Del Bimbo PPA (2010) 3d face recognition using isogeodesic stripes. *Pattern Anal Mach Intell* 32:2162–2177
100. Ballihi L, Amor MDASDABB (2012) Boosting 3-d-geometric features for efficient face recognition and gender classification. *Forensics Secur* 7:1766–1779
101. Mpiiperis I, Malassiotis MGSS (2008) Bilinear models for 3-d face and facial expression recognition. *Inform Forensics Secur* 3:498–511
102. Amberg B, Knothe R, Vetter T (2008) Expression invariant 3d face recognition with a morphable model. In: FG'08 7:1766
103. Al-Osaimi F, Bennamoun M, Mian A (2009) An expression deformation approach to non-rigid 3D face recognition. *Int J Comput Vis* 81(3):302–16
104. ter Haar RC, Velkamp F (2010) Expression modeling for expression-invariant face recognition. *Comput Graph* 34:231–241
105. Alyuz N, Gokberk B, Akarun L (2008) A 3d face recognition system for expression and occlusion invariance. In: Biometrics: theory, applications and systems
106. Mayo M, Zhang E (2009) 3d face recognition using multiview key point matching. *Advanced video and signal based surveillance*. pp 290–295
107. Alyuz N, Gokberk LSRVLAB (2012) Robust 3d face recognition in the presence of realistic occlusions. *Biometrics (ICB)*:111–118
108. Alyuz N, Gokberk B, Akarun L (2013) 3-d face recognition under occlusion using masked projection. *Inform Forensics Secur* 8:789–802
109. Hariri W, Tabia H, Farah N, Benouareth A, Declercq D (2017) 3d facial expression recognition using kernel methods on riemannian manifold. *Eng Appl Artif Intell* 64:25–32
110. Zeng Z, Pantic GRTHM (2009) A survey of affect recognition methods: audio. *Patt Anal Mach Intell* 31:39–58
111. Nicolaou M, Gunes MPH (2011) Continuous prediction of spontaneous affect from multiple cues and modalities in valencearousal space. *Affect Comput* 2:92–105
112. Vinciarelli A, Pantic DHCPIPFDMMSM (2012) Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *Forensics* 3:69–87
113. Kemelmacher-Shlizerman I, Basri R (2011) 3d face reconstruction from a single image using a single reference face shape. *Pattern Anal Mach Intell* 33:394–405
114. Park U, Tong AKJY (2010) Age invariant face recognition. *Pattern Anal Mach Intell* 32:947–954
115. Lanitis A, Taylor CJ (2000) Robust face recognition using automaticage normalization. In: Mediterranean electro-technical conference Vol.2, pp 478–481
116. Lanitis A, Taylor TCC (2002) Toward automatic simulation of aging effects on face images. *Pattern Anal Mach Intell* 24:442–455
117. Boussaad L, Benmohammed M, Benzid R (2016) Age invariant face recognition based on dct feature extraction and kernel fisher analysis. *J Inform Process Syst* 12(3):392–409

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
