IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 March 2023, accepted 16 April 2023, date of publication 26 April 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270713

RESEARCH ARTICLE

Face ShapeNets for 3D Face Recognition

MARWA JABBERI[1,2], (Member, IEEE), ALI WALI[2], (Senior Member, IEEE), BILEL NEJI[3], TAHA BEYROUTHY[3], AND ADEL M. ALIMI[2,4], (Senior Member, IEEE)

[1]ISITCom, University of Sousse, Sousse 4011, Tunisia
[2]REsearch Groups in Intelligent Machines (REGIM Laboratory), National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia
[3]College of Engineering and Technology, American University of the Middle East (AUM), Egaila 54200, Kuwait
[4]Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2006, South Africa

Corresponding author: Marwa Jabberi (marwa.jabberi@ieee.org)

This work was supported by the Tunisian Ministry of Higher Education and Scientific Research under Grant LR11ES48.

ABSTRACT In this paper, we present a deep learning-based method for 3D face recognition. Unlike some previous works, our process does not rely on face representation methods as a proxy step to be accepted by Convolutional Neural Networks (CNNs). Applying 2D CNNs to irregular domains such as 3D meshes is challenging. Therefore, we employed 3D ShapeNets to recognize faces covering the full 3D shape since 3D face datasets are available and 3D data augmentation techniques to enlarge 3D datasets are widespread. The reduced size of 3D datasets is overcome by an appropriate 3D data augmentation to train our model. 3D ShapeNets are commonly used to recognize and analyse objects. To the best of the authors' knowledge, this is the first time they are used for face recognition. This research work focuses on the preprocessing step. Whatever the nature of the face image is (either 2D or 3D representation), and whatever the acquisition conditions are, a 3D regular mesh of each input face image is first generated. Furthermore, meshes are converted to voxels to get the occupancy grid across all possible views. Finally, 3D ShapeNets are trained and recognition tests are performed. Indeed, 3D ShapeNets prove the efficiency and efficacy of 3D shape analysis task in 3D face recognition. The experimental results show that 3D face recognition using deep 3D CNNs such as 3D ShapeNets leads to significant improvement over the state-of-the-art performance on LFPW, BU3DFE, and FRAV3D datasets, with competitive recognition rates of 94.25%, 97.9% and 98.31% respectively.

INDEX TERMS 3D data augmentation, 3D face recognition, 3D meshes, 3D ShapeNets, feature extraction, voxelization.

I. INTRODUCTION

Security and anti-terrorism activities have raised great interest in biometric technology. Traditional methods, such as passwords and signatures, are therefore insufficient nowadays. In fact, biometric recognition systems have been extensively deployed worldwide in law enforcement, governments, and consumer applications [1], [2], [3], [4], [5], [6], [7]. In the last few years, face recognition (FR) methods have become the prominent biometric technique for individual authentication and verification. The face has sophisticated characteristics that are measurable, non-intrusive and contactless. FR is widely used in many domains, such as the military field, computer vision (CV) community, public security, access control, and daily life. It is a popular topic accepted by users.

2D FR methods give good performance. However, this representation has major problems, such as variation of lighting conditions, pose changes, and occlusions, etc. [8], [9], [10], [11], [12], [13], [14], [15]. To overcome the lighting conditions problems, researchers have migrated to other 2D representations and additional processing operations; however, this challenge remains an obstacle that degrades the recognition rates. In the case of pose variations, a face alignment step where all faces are in frontal poses is added. This leads to

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

an improvement in the recognition rate, but information loss is always found once the faces are passed through another representation or are modified by pose correction.

Therfore, 3D FR presents a good modality that overcomes the difficulties associated with pose changes and illumination conditions [16]. Indeed, the 3D domain opens up new horizons for enhancing the reliability of face-based identification systems. 3D shapes as geometric representations are considered one of the most important representations in object recognition, including faces [17].

Many theories about 3D representation and reconstruction are available [18]. Moreover, 3D-based methods are booming and consequetly 3D CNNs are currently in full development. A 3D CNN aims at representing geometric 3D shape, recognizing object category - in our case the class to which the face belongs-, completing full 3D shape, and predicting the next-best-view if the initial recognition is uncertain. Following the bibliographic research we performed, 2D face datasets are publicly available but 3D face datasets are hard to acquire. Scans and subjects are also limited and not sufficient for training a CNN.

The aim of this paper is to recognize 3D faces from an unconstrained face image using 3D ShapeNets as a 3D CNN since most of the current approaches do not fully exploit 3D informations. We focus on several main steps as following: 3D face datasets enhancement, 3D face representation accepted by 3D ShapeNets without domain reducing, and face verification for FR task.

Our main contributions are:

- 3D face dataset reconstruction in case of using 2D benchmarks since 3D face datasets are limited and not enough to be trained in a CNN.
- View-based 3D data augmentation since the out-of-plane rotation of the face does not cause a problem in 3D domain comparing with 2D face images.
- 3D face mesh processing and adapting to feed 3D ShapeNets.

The remainder of this paper is structured as follows: A literature review about the existing approaches of 3D FR, including a brief description on the objectives of our proposed method is presented in Section II. Section III provides an overview of our groundwork with justification of each step during this research. Details about the proposed work are presented using block diagrams and related algorithms. Section IV elaborates the experimental qualitative and quantitative results, and presents a comparison with a detailed analysis followed by a deep discussion on how to improve the efficiency and robustness of the proposed work. In Section V, we present the main conclusion and perspectives of the paper.

II. RELATED WORKS

In this paper, we are interested in 3D FR approaches since 2D FR has shown some drawbacks as previously cited. 3D FR rates exceed those of 2D FR rates due to 3D geometric representation which cover all the face regardless of the capture in the wild. 3D FR approaches can be classified as shown in Fig.1.

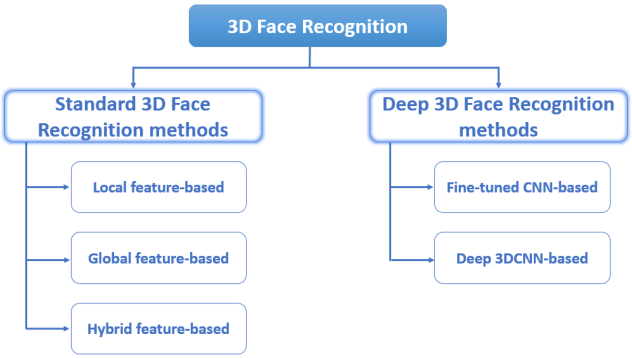


FIGURE 1. 3D face recognition categorization methods.

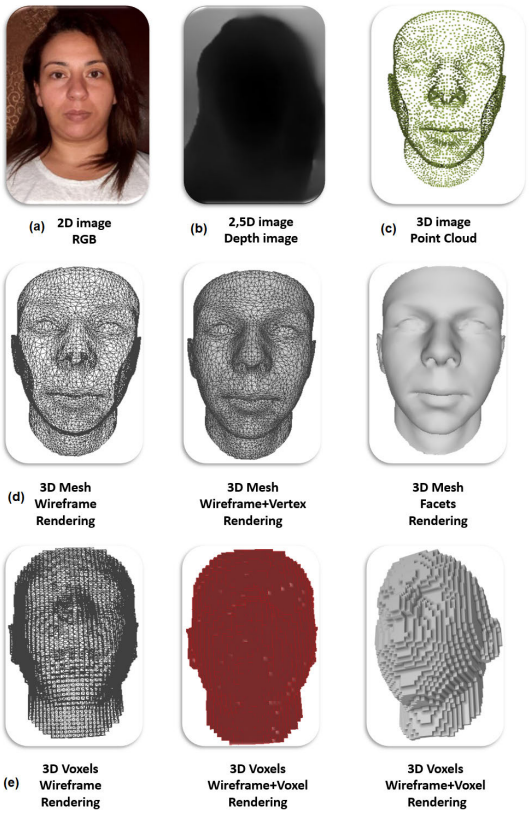


FIGURE 2. The most widespread 3D image representations.

A. OVERVIEW OF 3D DATA REPRESENTATIONS

Before studying 3D FR existing methods, we present the different 3D representations and rendering. Indeed, there are several representations; however, we will only present he most widespread [80].

- **Point Cloud:** This representation (Fig.2 (c)) is a set of points in the space. Each point contains its set of Cartesian coordinates (x, y, and z) that define the point

position. The point cloud represents the envelope 3D shape or object. It provides simple 3D data representations that make deriving meshes very easy using an appropriate surface reconstruction technique.

- **Mesh:** This representation (Fig.2 (d)) includes a collection of vertices (or vertexes), edges and facets which define a 3D shape or object. Meshes can be presented by triangles (triangular mesh), quadrilaterals (quads), or convex polygons (n-gons). Triangular meshes are the most common in Reality Capture workflows since they are the simplest. Vertices present x, y and z Cartesian coordinates. Facets include connectivities between different vertices. In the case of a triangular mesh, a facet requires to have exactly 3 vertices. An edge connects 2 vertices. Transformations in 3D meshes are quick and easy but they depend on high memory requirements.
- **Voxel:** This representation (Fig.2 (e)) is based on cubical units that can be used to represent 3D models. A voxel looks like a pixel in a 2D representation (bitmap) (Fig.2 (a)). Voxels are encoded with their values, which are space occupancy. A 3D object or shape modelized as voxels-based model is a discretization of "3D pixels."

Furthermore, depth images, also named RGB-D images, are in 2.5D as shown in (Fig.2 (b)). Depth images are technically a 2D representation since they are a construction of an apparently three-dimensional environment from 2D retinal projections.

Each 3D representation has its unique characteristics that can be used in a specific application. It is true that this representation is expensive with regard to acquisition devices, processing, analysis, and manipulation. However, it allows to overcome the lighting and pose variations problems of 2D representation.

B. STANDARD 3D FACE RECOGNITION APPROACHES

Methods classified into this category are divided into three classes: local feature-based, global or holistic feature-based and hybrid feature-based methods. In this category, traditional techniques and algorithms are used for facial features extraction and selection. The conventional pipeline is shown in Fig.3.

In the Offline phase, the 3D faces dataset follows the processing steps. Smoothing filters are used for denoising. In addition, Delauay triagulation is employed for holes filling. Finally, segmentation is used to extract salient part of the face with the purpose of ignoring the hair, neck, ears, etc. The feature extraction step is carried out to compute the biometric signature of each face. These signatures are later stored.

The Online phase is carried out at each interrogation of the dataset by the user, where a query face goes through the same steps, i.e. preprocessing and feature extraction. Furthermore, a feature matching is established to compare and analyze the query and the saved signatures.

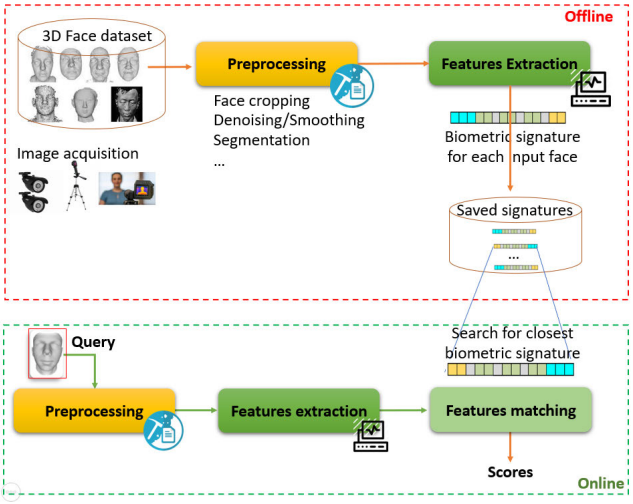


FIGURE 3. Overview of standard 3D face recognition approaches pipeline.

1) LOCAL FEATURE-BASED 3D FACE RECOGNITION APPROACHES

In this category, local features are extracted from parts of the face, i.e. descriptors are generated to characterize significant parts (patches) of the facial surface. Most of the related recent research is mainly focused on establishing face segmentation or salient regions extraction in the preprocessing phase to identify the significant parts of the face. Local features are generated from discriminative patches of the face, such as the nose, eyes, and mouth. Table 1 presents a study of some recent local feature-based approaches.

TABLE 1. Local feature-based 3D face recognition approaches.

Year /Author	Used Technique	Used Descriptors	Results (%)
Hariri et al. [19] (2016)	Local surface computing based on covariance descriptors	Feature points from covariance descriptors	99.2 97.81
Deng et al. [20] (2017)	Intrinsic correlation of local covariance descriptor and Riemannian kernel sparse coding	MeshDOG Keypoints Local GH descriptors	97.3
Soltanpour et al. [21] (2019)	Normal maps from local derivative variation	LDP descriptors Histogram construction	99.3 95.36
Shi et al. [22] (2020)	Local surfacecomputing based on LBP	LBP descriptors	96.83
A. Taleb et al. [79] (2022)	High order local descriptors for feature extraction for sparse representation	Multiscale depth LDP SIFT points	93.46 97.76

2) GLOBAL FEATURE-BASED 3D FACE RECOGNITION APPROACHES

The global or holistic feature-based approaches for 3D FR extract features from the entire facial surface. In table 2, several recent approaches are presented. The used descriptors are not dedicated to give information on a specific part or region of the face. They use the entire facial surface features.

TABLE 2. Global feature-based 3D face recognition approaches.

Year/ Author	Used Technique	Used Descriptors	Results (%)
Yue Ming [23] (2014)	Estimating the regression matrix to decribe intrinsic facial surface features	Curvature, orthogonality and spectral regression informations	93.95
Tang et al. [24] (2015)	Construction of asymptotic cones based principal curvature	Local Normal Patterns descriptor	93.16
Gilani et al. [25] (2017)	Point-to-point correspondence	Deep Landmark Identification Network	98.1
Neto et al. [26] (2017)	Feature extraction from depth map	3DLBP descriptor	89.08
Peter et al. [27] (2019)	Non linear kernal approach	Kernel based PCA	-

Global 3D FR methods are time consuming, but very efficient to synthesize the complete face.

3) HYBRID FEATURE-BASED 3D FACE RECOGNITION APPROACHES

Hybrid or multi-modal approaches are a blend of the global and local techniques. They are the most commonly used in pattern recognition. Global features serve to get features of the general appearance or shape of the face. However, local features serve for strengthening of the biometric signature of the face and they focus on certain critical regions such as the mimic regions.

Indeed, hybrid methods take the advantages of both the global and local methods by combining the detection of geometric characteristics with the extraction of local salient facial characteristics. This mixture increases the recognition rate despite the high complexity. In Table 3, we present some methods belonging to this category.

C. DEEP 3D FACE RECOGNITION METHODS

Following the emergence of new acquisition techniques, 3D devices and 3D modeling tools, 3D shapes have become simpler and more understandable.

CNNs are also sophisticated means for analysis, features extraction, classification and 3D shape verification. Since the power of deep learning models has allowed researchers to build deep generative models for 2D shapes, deep learning models for 3D shapes are also possible. In this classification, there are two categories: Fine-tuned 2D CNNs-based and Deep 3D CNNS-based approaches.

1) FINE-TUNING CNN-BASED APPROACHES

In this category, Deep 2D CNNs are used for 3D face analysis, feature extraction, and classification. In Table 4, selected approaches are presented. There are two types of solutions:

- 3D domain representation reducing and passing from 3D to 2D or 2.5D (depth) representations.
- Adaptation of convolutional layers channels to acquire 3D inputs.

TABLE 3. Hybrid feature-based 3D face recognition approaches.

Year/Author	Used Technique	Used Descriptors	Results (%)
Bagchi et al. [28] (2014)	Face surface registration and restoration	ICP and PCA	91.3
Elaiwat et al. [29] (2015)	Identifying keypoints in textured 3D face surface and texture features extraction	Keypoints coeficients Digital curvelet transform	97.1 95.01 91
Liang et al. [30] (2017)	Facial landmarks model-based search and face structure-based analysis	HK curvature	94.79
Shi et al. [22] (2020)	Depth face feature extraction	LBP Histogram normalization	96.83
Dutta et al. [31] (2021)	Complement component conjunction for data level fusion to generate vast face space	Complement componenthy bridization based on Genetic algorithm based feature-level fusionand swarm optimization	97.86 98.25 99.89

TABLE 4. Fine-tuning CNNs-based approaches.

Year/Author	Used Technique	Model	Rates (%)
Kim et al. [32] (2017)	Depth map generation from 3D scans	VGG-Face (pre-trained from 2D faces) Pre-trained LS-SRC Deep normal pattern combined with VGG-Face	99.2 95 87.7 98.01 96.13 97.60
Li et al. [33] (2017)	3D facial surface projection to 2D plane Location-sensitive space representation		
Luo et al. [34] (2019)	Depth map generation from 3D point cloud, 2D, and 3D data fusion	9-Layer CNN	-
Cai et al. [35] (2019)	3D data projection to range images Scale normalization with only 3 facial landmarks	Pre-trained ResNet	100 99.75 98.92 96.73
Dutta et al. [36] (2020)	Depth map generation from 3D scans	SpPCANet (learned filter bank)	95.87 96.93 88.59 88.80 98.59 98.54
You et al. [37] (2020)	'flatten' geometric information with the color from 3D to 2D plane	Pre-trained VGG-Face Fine-tuned with generated multi-channel face image	98.6 98.6

Fine-tuning CNN-based approaches consist in reducing the 3D data into another representation such as, 2D plane projection (RGB rendring), generation of depth images (depth maps), and range images creation.

The advantage is summarized in the use of deep 2D CNNs, like VGG-Face or ResNet, which are pretrained from large

TABLE 5. The most widespread data augmentation techniques for 3D faces.

Year/Author	Data Augmentation Technique
Kim et al. [32] (2017)	Facial expressions modification by generating random patches from 3D faces. Facial expressions are transfered by adding expression to the multi-linear 3DMM [38], [39]. Pose variation.
Luo et al. [34] (2019)	Merging 2D face color maps and facial depth maps.
Cai et al. [35] (2019)	Pose, transformational, and resolution variantions.
Dutta et al. [36] (2020)	Translation and rotation.
Marwa et al. [83] (2023)	Generative, data augmentation based on resolution enhancement accross different poses.

2D face images. This makes features extraction very efficient because deep 3D CNNs suffer from lack of 3D datasets. Yet, the disadvantage lies in the passage of the 3D representation, which is very close to reality in terms of volume, to the 2.5D or 2D representation. This conversion or transformation generates data loss which subsequently decreases the recognition rates.

Fine-tuning CNN-based approaches were successfully used on data augmentation to overcome the problem of limited 3D datasets. Several data augmentation techniques for the used data are listed in table 5.

2) DEEP 3D CNN-BASED APPROACHES

In this category, 3D scans are used directly without the conversion to other representations or planes. Deep 3D CNNs work with 3D faces (meshes, point clouds, voxels, etc). 3D CNNs can be classified as following:

Voxel-based deep 3D CNNs: This classification involves the volume and grid occupancy. In fact, a 3D shape is represented as voxels and a 3D CNN is applied over the entire 3D volume. O-CNN [40] and 3D ShapeNets [41] are the most known voxel-based deep 3D CNNs. Since most 3D face datasets are in the form of a triangular mesh, a voxelisation step is required. The success of the voxelization stage depends on the starting mesh which must be a regular one. It is worth noting that the previously mentioned deep 3D CNNs are used for 3D objects recognition, completion, and prediction.

Mesh-based deep 3D CNNs: In this classification, deep CNNs use meshes and graphs [42]. Authors in [43] worked on the generalizations of deep CNNs on non-Euclidean domains to save significant parts of the object or its fine details and topological structures. In [44], deep CNNs assume the input shape as a mesh with a sphere-like topology where the goal was to introduce a conformal mapping from the mesh surface into a flat torus. The main shortcoming of these methods is that they significantly limit the exploitation of the maximum possible workspace available, especially as the neighborhood factor is taken into account since the mesh or graph represents a set of relationships between points in space.

3D deep CNNs based on point cloud: This category of deep 3D CNNs is inspired from the voxel-based deep 3D CNNs, including VoxNet and 3D ShapeNet with a small change. It involves the use of density occupancy grids representations for the input data. In [45], the authors proposed PointNet (deep 3D CNN) based on the implemented architecture of the Point Cloud Library (PCL) [46] for 3D point cloud processing. To the best of our knowledge, the results showed that PointNet is the fastest deep 3D CNN.

There are deep 3D CNNs that are suitable for extracting spatial and temporal features in facial video frames [47], [48], [49], [50]. In this study, we are interested in deep 3D CNNs that extract the geometric features of a 3D face. 3D FR using deep 3D CNNs is still developing. The major problem lies in the limitation of the 3D face datasets. Indeed, training a deep CNN requires a large dataset. On the other hand, due to our state of the art study, several 3D CNNs have been used in the context of object recognition.

In this paper, the objective is to recognize 3D faces using deep 3D CNNs. 3D scans of faces are used without reducing the domain of representation. For this reason, we used a 3D CNN working directly with this complex geometric representation. The advantage of this contribution is the use of 3D faces as they are defined, but the major problem remains with the limited size of the current datasets. Therefore, in order to outperform deep 2D face recognition, a big dataset for 3D CNN training is needed.

III. 3D FACE RECOGNITION BASED ON 3D ShapeNets

A. OVERALL PIPELINE

Our method consists of 3 main phases: 3D Data preprocessing, 3D Data augmentation, and features extraction and classification based on 3D ShapeNets which is a 3D CNN. Our contribution consists in applying 3D ShapeNets to 3D FR, which has not previously been used to the best of our knowledge.

It is also important to mention that in our work we did not make domain reducing of 3D data representation because we used a deep 3D CNN which had as input 3D volumes. One of the goals of this paper is to find a solution to the issue related to the small number of subjects in 3D face datasets that 3D CNNs can validate. We used two types of input:

- 3D reconstructed faces from 2D face images since 2D face datasets are sophisticated and have a large number of classes, each having a large number of subjects. It is true that 3D reconstruction from single 2D image is not as one captured with 3D acquisition devices, but we gained a 3D representation that was mathematically valid and very close to the real world.
- 3D face datasets that were acquired by 3D devices. Certainly, we have applied preprocessing and data augmentation steps to enlarge the number of 3D scans.

It is worth mentioning that this stage is time-consuming; however, as it was a part of our previous work, we opted for using it [51]. Furthermore, we thought of the good quality of the mesh built once the initial one was obtained in case of using 2D datasets. We focused on mesh regularity to facilitate the voxelization step. We used butterfly algorithm for mesh subdivision accompanied by BPA algorithm for good quality and face shape preservation. Fig. 4 presents the proposed overall pipeline of our work. Each step will be detailed in the following subsections.

B. DATA PREPROCESSING

1) 3D FACE RECONSTRUCTION

To start with, a comparative study between 2D and 3D datasets is performed. It was concluded that the 2D datasets were wider and contain a big variety of classes as shown in table 6.

TABLE 6. Study of the most widespread datasets of 2D and 3D faces.

2D datasets			
Name	**#Classes**	**#Faces**	**Challenges**
AFW [52]	205	468	Aging Occlusions Expressions Poses
LFW [53]	5749	13233	Lighting Poses Resolution Occlusions
CASIA -Web Face [54]	10575	494414	Poses
AFLW [55]	-	5993	Poses Expressions Ethnicity Age Gender
LFPW [56]	-	1432	Poses Occlusions
Multi -PIE [57]	337	750000	Lighting Poses
FRAV2D [58]	109	3488	Poses Lighting Occlusions
3D Datasets			
Name	**#Classes**	**#Images**	**Challenges**
GavabBD [59]	61	549	Poses Expressions
Texas 3D [60]	105	1149	Gender Ethnicity Expressions
Bosphorus [61]	105	652	Poses Occlusions Expressions
BU-3DFE [62]	100	7500	Expressions
3D TEC [63]	214	856	Expressions
FRGC.v2 [64]	466	28049	Expressions Age
FRAV3D [65]	106	1696	Poses Lighting

3D datasets are not large enough to train a deep CNN. We treat two types of input: faces coming from 2D datasets (i. e. RGB images) and 3D faces captured by 3D devices (i.e. 3D scan files such as ".off"). Therefore, we establish 3D face reconstruction from faces in 2D datasets since our objective is to extract features using deep 3D CNN which is 3D ShapeNets and the input should be a 3D volume applied to perform 3D face recognition. In a previous work [51], we proposed 3D face reconstruction in the context of face alignment and frontalization. This implies that 3D reconstruction may be used to deal with several 2D FR problems and limitations.

In Pseudo Code (1), we present the generalized steps of 3D reconstruction (details are presented in our previous work [51]).

Pseudo Code (1): 3D Mesh Reconstruction From 2D Face Image

Input: 2D face images
Output: 3DFace
1. Read input images from dataset
2. For #image
  2DFace ← Face Detection and Cropping from input images
  2D$_{Landmarks}$ ← Landmarks Detection and Extraction (2DFace)
  68$_{Landmarks}$ ← Estimate the location of 68 (x, y)-coordinates that map the facial structures on the face (2DFace) using dlib library
  Edges$_{Canny}$ ← Edge Detection (2DFace, 'Canny')
  Edges$_{Prewitt}$ ← Edge Detection (2DFace, 'Prewitt')
  Regions ← MSER Features Detection (2DFace)
  2D$_{keypoints}$ ← Reshape (68$_{Landmarks}$, Edges$_{Canny}$, Edges$_{Prewitt}$, Regions)
  3D$_{keypoints}$ ← Delaunay Triangulation (2D$_{keypoints}$)
  3DFace ← 3D Mesh Generation (3Dkeypoints)
  Write off (3DFace)

2) FACIAL SURFACE EXTRACTION

At this stage, each 3D face follows some pretreatment steps. Whatever the nature of the 3D face is, preprocessing is essential for meshes generated from 2D images or meshes acquired directly by 3D cameras. Preprocessing is conducted for two reasons. First, it is conducted for facial surface extraction. It ignores the useless information, namely the background, ears, neck, etc. Secondly, it is performed for remeshing which is essential and it served in the voxelization step, which will be detailed later.

For facial surface extraction (Fig.5), we start by detecting the nose tip (Pseudo Code (2), Line 3), which is our starting point. We represent it by the point having the largest depth value along the Z axis following this Pseudo Code.

We determine the extraction radius suitable for any face shape (r = length of Bounding Box * 0.6) (Pseudo Code (2), Line 7). It is chosen following a series of experiments and the extracted area is calculated using the geodesic distance to obtain an oval shape compatible with any facial shape. It is worth noting that the extraction method is invariant to scale and pose variations.

For 3D faces generated from 2D images and undergoing pose variations, we notice the presence of missing parts,

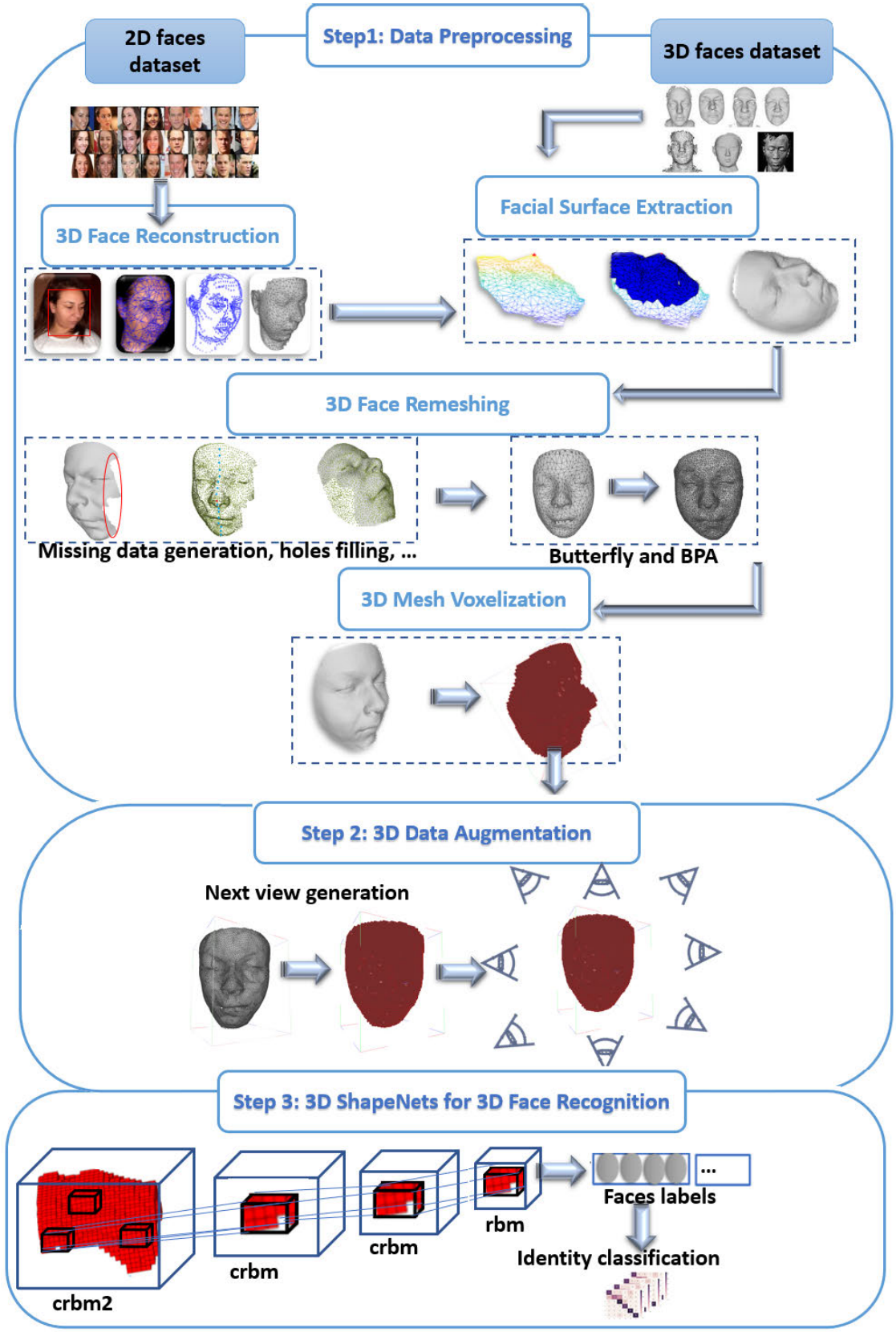


FIGURE 4. Overview of our proposed method.

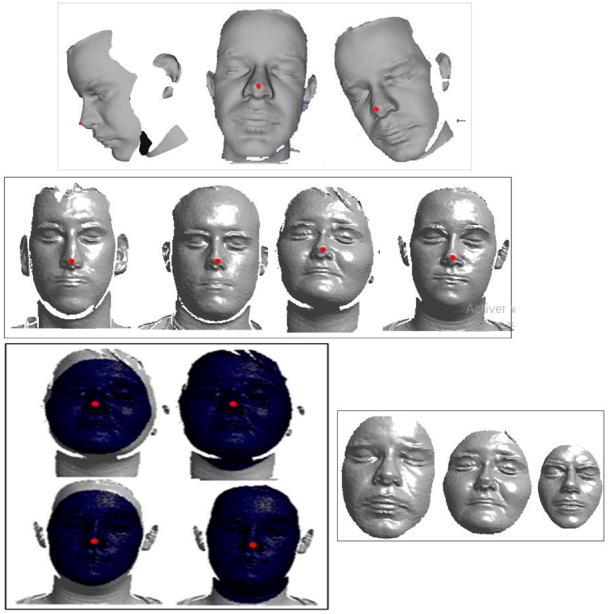


FIGURE 5. Facial surface extraction using region growing segmentation method.

Pseudo Code (2): Salient Facial Surface Extraction

Input: 3D Face
Output: $3DFace_{Patch}$

1) [vertex, face] ← Read 3DFACE
2) [m, i] ← max (vertex (:, 3))
3) $Nose_{Tip}$ ← vertex (i,:)
4) A ← triangulation2adjacency (face)
5) Box, LengthBB, r ← CalculateBoundingBox (vertex)
6) vring ← ComputeVertexRing (face)
7) $3DFACE_{Patch}$ ← RegionGrowingGeodesic (face, vertex, vring, i, r)
8) Write (['$3DFace_{Patch}$.off'], $3DFace_{Patch}$.vertex, $3DFace_{Patch}$.face)

called self occlusions. Therefore, correction consisting in missing parts reconstruction and meshing is made. This correction is detailed in our previous research work [51].

After 3D face shape correction and meshing missing parts (Fig.6), the resulting 3D face is ready for the remeshing step. First of all, this step is used to connect the new added vertices. In addition, the facial surface subdivisions of the mesh is performed to properly prepare our 3D mesh for voxelization.

3) FACE REMESHING

This step is performed using the Butterfly subdivision algorithm [66] and the Ball Pivoting Algorithm (BPA) [67] for 3D mesh subdivision and triangular interpolation as shown in Fig. 7.

This process is very essential for 3D mesh preprocessing. We check mesh regularity by computing the vring of each vertex, which is the connection number of each one. To get a regular mesh, all vertices must have the same vring.

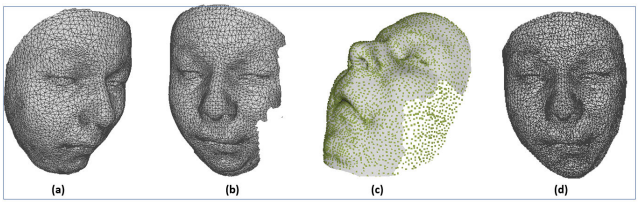


FIGURE 6. Missing parts reconstruction and meshing: (a) Reconstructed 3D face from 2D face image, (b) Face rotation to show missing parts caused by pose variation, (c) Symmetric vertices generation, (d) Shape correction and missing parts reconstruction.

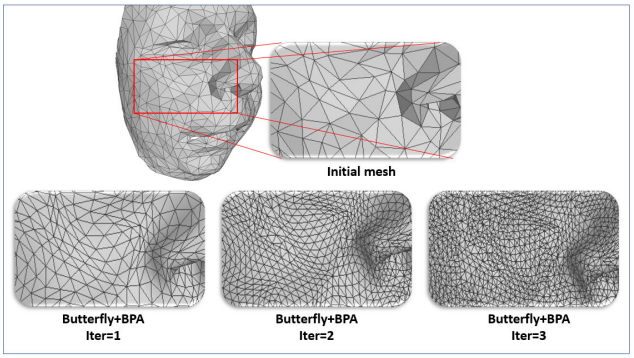


FIGURE 7. 3D mesh subdivision based on butterfly and BPA with shape preserving.

Therefore, at each subdivision iteration, vertices and connections are added. Then, a regularity test is performed if the criterion is not checked. BPA operates refinement, interpolation and shape preserving. Indeed, these two algorithms go well together and guarantee a very good remeshing.

4) MESH VOXELIZATION

In this work, we used 3D ShapeNet, which is a voxel-based deep 3D CNN. For this reason, the voxelization step was performed. It consists in representing the obtained 3D mesh as a binary tensor in the following way:

- '1' indicates that the voxel is inside the mesh surface.
- '0' indicates that the voxel is outside the mesh or in an empty space.

As long as mesh regularity is checked, the 3D voxed volume will be of very good quality and each vertex in place will be represented by a voxel. It is true that we cannot see the details of the facial surface in this representation. However, each voxel is significant in the occupancy grid (Fig. 8). This is why 3D ShapeNets were designed for objects and not for complex surfaces such as faces. In this work, we reveal this challenge to see if this 3D CNN is really effective for facial recognition.

C. 3D DATA AUGMENTATION

As detailed in the previous sections, 3D data augmentation is essential to train a deep CNN, especially when using 3D datasets. Since we converted 3D meshes into voxels, we obtained an occupancy grid. Indeed, we put the 3D face

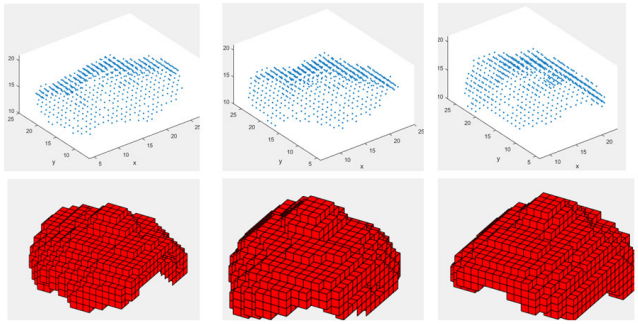


FIGURE 8. 3D mesh voxelization.

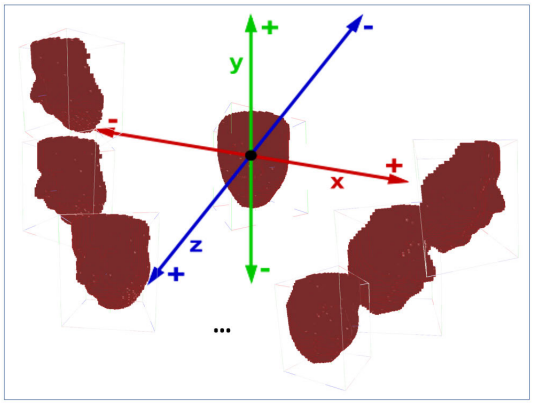


FIGURE 9. View-based 3D data augmentation: generation of 12 views.

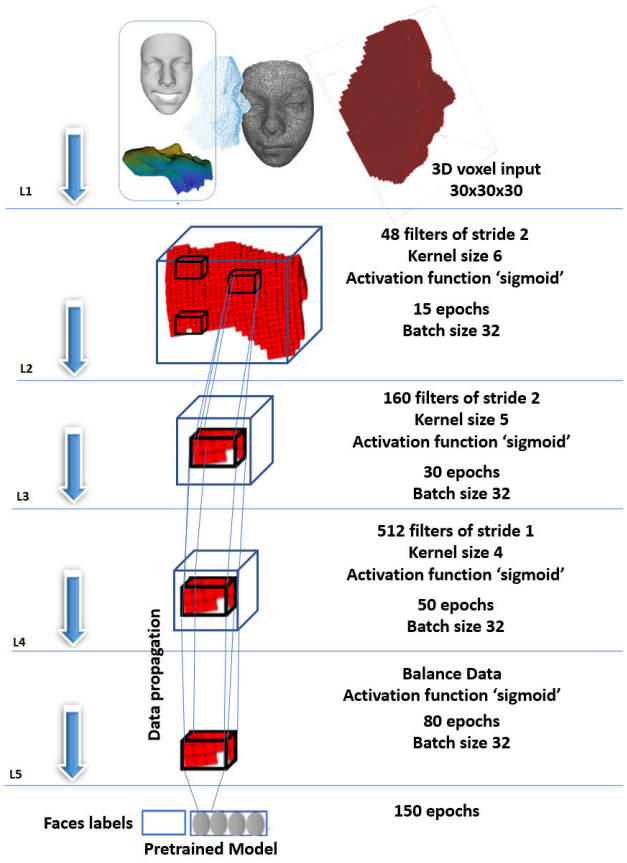


FIGURE 10. 3D Face ShapeNet architecture.

volume in a box that we can look at from different angles. It is not a pose variation because our volume is stable in the box, but it is a change of angles of view. For this reason, it is called a view-based 3D data augmentation technique. As a result, different views give meanings to space occupation. For each face, we generated 12 views equivalent to a consecutive rotation of 15° accross yaw (180°: 15° = 12 views), as illustrated in the following figure.

The choice of the number of views is made following a series of experiments, which allowed us to notice that for each 15°, there are remarkable changes in the occupancy grid with an important meaning. Contrary to objects recognition, the angle of rotation can be reduced to less than 15° over 360° to get better results.

D. 3D ShapeNets FOR 3D FACE RECOGNITION

In this work, we used 3D ShapeNets as a 3D CNN which takes as input a 3D volume. We no longer need to reduce the representation of our input data or passing through fine-tuned CNNs.

As mentioned in the previous subsections, we represent 3D faces as a probability distribution of binary variables on a 3D voxel grid using a Convolutional Deep Belief Network (CDBN) [68]. CDBNs are hierarchical generative models for deep learning, which are effective in image processing and object recognition. They contain multiple layers of Convolutional Restricted Boltzmann Machines (CRBM) [78]. CDBNs work comfortably with high-dimensional images including complex shapes representations and are translation-invariant. The 3D ShapeNet architecture is represented in Fig.10.

1) 3D FACE ShapeNets ARCHITECTURE

The first layer: (L1) is the input layer. The grid size in our experiments is 30 × 30 × 30, which is equivalent to 165 × 165 2D image. Face labels are presented as standard 1:K softmax variables and they are duplicated for the final RBM in order to enforce the label training.

The second layer: (L2) takes as input a volume pointing to the 3D face mesh with 48 filters of size 6 and stride 2. The volume is represented as a probability distribution of binary variables (1 refers to the occupied space and 0 for the empty space in the occupancy grid). Only non-empty learning signals are considered to ignore training distraction and to get meaningful features.

The third layer: (L3) has 160 filters of size 5 and stride 2. Each filter has 48 × 5 × 5 × 5 parameters. The convolution layer is then used to reduce model parameters by weight sharing while ignoring pooling in the hidden layers whose purpose is to prohibit properties invariance for the recognition task. In this layer, the CRBM training is used

to establish different convolution CUDA kernels for parallel computational elements with the next layers. We notice that the purpose of RBM in 2D CNNs consists in learning centered filters and departing from boundaries. In 3D ShapeNets, RBM is used to collect only persistant learning signals.

The fourth layer: (L4) has 512 filters of size 4 and each CRBM filter is connected to all the feature channels in the previous layer.

The fifth layer: (L5) is a standard fully connected RBM with 1200 hidden units. Data are propagated from the bottom.

The last layer: (L6) has 4000 hidden units and takes as input stacked variables and parameters which are multinomial label variables and Bernoulli feature variables. The purpose is to model the set of possible outcomes, model evaluation, and training costs. The final hidden activations are computed using Fast Persistent Contrastive Divergence (FPCD) [76] and RBM for the joint training of data and labels.

2) 3D FACE ShapeNets TRAINING

The first layers are trained using a different objective function called Contrastive Divergence (CD), which minimizes the divergence between data distribution and derivatives with regard to the parameters that can be accurately and efficiently approximated [69].

$$CD_n = KL(p_0 \parallel p_\infty) - KL(p_n \parallel p_\infty) \tag{1}$$

where $\mathbf{p}_0$ is the data distribution, and **KL** is the Kullback-Leibler divergence. They are computed as follows:

$$KL(p_0 \parallel p_\infty) = \sum_x p_0(x) \log \frac{p_0(x)}{p(x; W)} \tag{2}$$

W refers to the probability distribution parameters over x. The purpose of CD (Eq. 1) consists in reducing computation per gradient step and the variance of estimated gradient. Mathematical demonstration is summarized in [77].

In contrast, the top layer is trained using FPCD. In the first layer, only non-empty voxels are taken into account. The purpose is to minimize the whole volume for RBM and to distract learning. Once the system is pretrained, we conduct 3D classification and retrieval experiments to evaluate extracted features along all previous layers. Other technical configurations and settings of 3D ShapeNet are detailed in [41].

3) 3D FACE ShapeNets CLASSIFICATION

3D ShapeNets have shown good results in object recognition and proved the model ability of features learning from 3D data automatically. In this work, we demonstrate that this model is also effective in 3D FR.

3D FR using 3D ShapeNets is conducted based on features evaluation where a linear SVM is trained for 3D face scans classification with regard to average classes accuracy for marching. Moreover, L_2 distance is established for pairwise similarity measurement. A ranked list of the remaining test data is returned in response to a query from the test set, using the similarity measure. Indeed, the recognition task consists in estimating the distribution between tested and trained features. The classification method is free energy.

IV. EXPERIMENTAL RESULTS

The conducted experiments are performed for LFPW, BU-3DFE, and FRAV3D datasets using 3D ShapeNets. The code implementation is based on the Graph toolbox, image Processing MATLAB toolbox for 3D mesh processing, and MeshLab which is linked to the NVIDIA packages to accelerate training. All our experiments were performed using NVIDIA CUDA development 9.2 and were run on intel (R) Core (TM) i7-7500U, 2.70 GHz and 2.90 GHz with 8 Go RAM.

For model pretraining, we set the following parameters: Labels are duplicated 10 times for the final RBM with the purpose of enforcing label training. Epochs and learning rate values are presented in Table 7. **CRBM 2** is the Convolutional RBM training of the second layer which uses different convolution CUDA kernels with L3 and L4. **CRBM** is the Convolutional RBM training of L3 and L4. **RBM** training with positive and negative phases uses persistant CD. **RBM Last** consisting of FPCD RBM training for the joint training of data uses labels sparsing of the last top layer.

TABLE 7. Model pretraining parameters.

Layer	Epochs	Learning rate	Batch size	Training Model
L2	15	0.01	32	CRBM 2
L3	30	0.01	32	CRBM
L4	50	0.01	32	CRBM
L5	80	0.003	32	RBM
L6	150	0.003	32	RBM Last

Number of epochs, batch size and learning rates were chosen after a series of experiments using optimizers [81], which resulted in finding the appropriate model hyperparameters. In addition, we have taken into consideration the fact that the model memorizes the data when there are too many epochs and fails to learn it. In this paper, the goal is to highlight that 3D ShapeNets can be applied on complex shapes like the face.

Once the model is pretrained with the full 3D shape, discriminative recognition test is established using a seperate test data. Predicted labels are estimated while calculating the free energy for each label hypothesis. We compute accuracy as follows:

$$Accuracy = \Sigma\left(\frac{predicted\ label == Label}{n}\right) \tag{3}$$

As an evaluation metric, we utilized the AROC and ROC curves which are used in such issues. AROC is an accomplishment plot of a classification model at each classification threshold calculated by numerically integrating an ROC curve. Therefore, we started with the calculation of ROC

curve as following:

$$\text{True Positive Rate} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})} \quad (4)$$

$$\text{False Positive Rate} = \frac{\text{False Positives}}{(\text{False Positives} + \text{True Negatives})} \quad (5)$$

In our experiments, we present the AROC curves by drawing the area under the ROC curve and positioning the accuracy found in each AROC.

A. EXPERIMENTATION AND RESULTS ON LFPW DATASET

LFPW dataset [56] is a collection of face photographs that are specifically devised to accurately understand and investigate unconstrained face recognition. LFPW contains 3,000 face images acquired from various sources from the World Wide Web. The associated persons' names were utilized to label every individual face in the dataset.

In this work, we used the outcomes of image sets from an evaluation of face parts (facial fiducial points) which were trained on 1,132 images and tested on 300 images. Therefore, the dataset was divided into two parts, one for training and another for testing.

To validate our work and the revealed challenge when using 3D ShapeNet for 3D face recognition, we have established 2 supplementary classifications for this database as shown in Table 8. Apart from the original decomposition, we classified train and test images manually according to gender-related (male or female) and age-related (children, adults, and elderly people).

TABLE 8. Proposed classification on LFPW dataset.

Classification		Before data augmentation		After data augmentation	
		Train	Test	Train	Test
Original classification		811	224	9732	2688
Gender-related classification	Male	459	113	5508	1356
	Female	352	111	4224	1332
Age-related classification	Children	13	5	156	60
	Adults	701	184	8412	2208
	Elderly	97	35	1164	420

It is important to mention that for each 2D face image, we generate a 3D face mesh. Once the mesh is created, we carried out preprocessing operations for facial surface extraction. A remeshing step is also performed to have a regular mesh. Indeed, having a regular mesh, allowed the voxelization step to be conducted perfectly, which increases the recognition rate.

Once the dataset is ready, we establish 3D view-based data augmentation. As a result, for each face volume, 12 views are created. We create the pretrained model using training volumes. The recognition test is performed using test volumes. Therefore, after running the pretrained model, we test the recognition ability (discriminative) of the model using a seperate test data. The classification method is free-energy. The obtained results are presented in the following table.

TABLE 9. Quantitative results on LFPW dataset.

Classification	Accuracy (%) 3D ShapeNet	Accuracy (%) VGG-16
Identities	94.25	89.23
Gender	86.68	82.79
Age	79.17	73.57

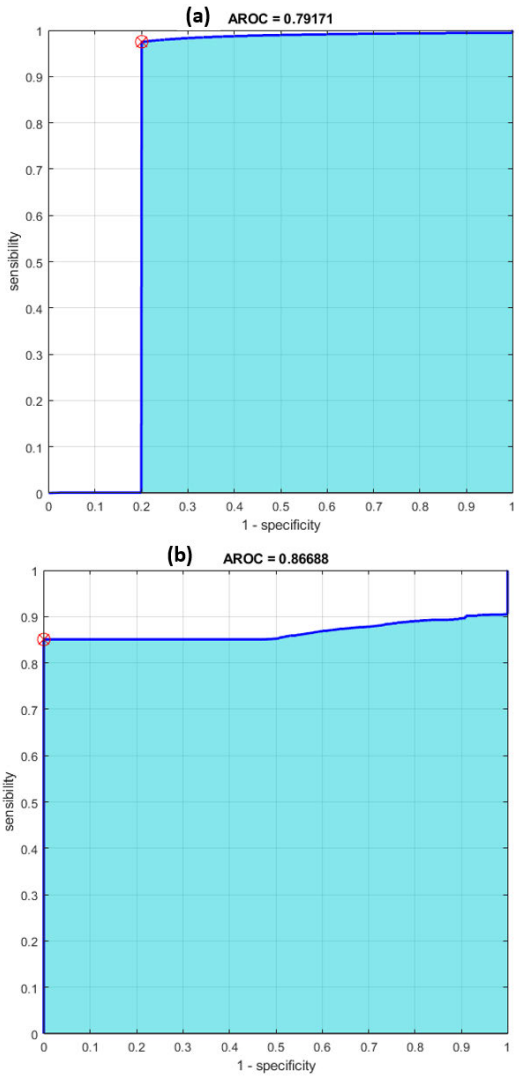


FIGURE 11. ARoc curves of LFPW dataset: (a) ARoc of age-related classification, (b) ARoc of gender-related classification.

The recognition rate is about **80%** for the age-related classification using 3D ShapeNet (Fig.11(a)). On the other hand, we got a recognition rate of approximately 73.57% using images that are trained and tested with VGG-16 [82], which is a 2D CNN. In the recognition testing for gender classification, we obtained around **87%** using 3D ShapeNet

(Fig.11(b)) and 82.79% using VGG-16. For face identities recognition, we obtained **94.25%** using 3D ShapeNet against 89.23% using VGG-16.

It is important to mention that for 2D recognition test using VGG-16, no additional pretreatment (such as alignment) was conducted on the faces, except for face detection and face cropping.

In the literature, LFPW dataset is used for face detection and facial landmarks localization. For this reason, we did not conduct comparisons with 3D FR additional peer approaches especially deep-based approaches on LFPW experiment. We first used this dataset to show that 3D reconstruction solves the problem of limited 3D datasets. In addition, the exploitation of 2D faces datasets in the framework of 3D FR is now possible even if it is from single 2D face image. Second, we would like to show that 3D FR rates are more important than 2D FR rates.

B. EXPERIMENTATION AND RESULTS ON BU3D-FE DATASET

This dataset [62] includes 100 subjects with 2,500 facial expression models. The database contains 100 subjects divided into 56 female classes and 44 male classes, with ages ranging from 18 to 70 years old, and having a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. Each subject performed seven expressions in front of a 3D face scanner. With the exception of the neutral expression, each of the six prototypic expressions (happiness, disgust, fear, anger, surprise, and sadness) includes four levels of intensity. Therefore, each subject presents 25 instant 3D expression models, resulting in a total of 2,500 3D facial expression models in the database. A corresponding facial texture image captured at two views is associated with each expression shape model. It is about $+45°$ and $-45°$. As a result, the database consists of 2,500 two-view texture images and 2,500 geometric shape models.

After establishing the preprocessing operations mentioned in the previous sections and the remeshing steps, we performed mesh voxelization. For each volume, we also generated 6 views as data augmentation. Therefore, the dataset is ready to be tested on 3D ShapeNets. We divided classes randomly into two parts, 70% for training and 30% for testing.

A recognition rate of **97%** was obtained with 3D ShapeNets (Fig.12) vs about 86% with VGG-16. We tested VGG-16 using the 2D images in the dataset. In table 10, a comparison with state of the art approaches is established.

C. EXPERIMENTATION AND RESULTS ON FRAV3D DATASET

FRAV3D database [65] includes 106 subjects, one woman to three men approximately, each having 16 pose variation. Apart from pose variations, there are expression and light variation, for each subject. 3D scans are captured using a Minolta VIVID 700 scanner. FRAV3D also contains texture information (2D images) and VRML files for 3D ones.

3D scans of this database are in poor condition due to the noise evoked by the acquisition devices. Therefore, we performed smoothing steps based on laplacian smoothing of noisy surface meshes [72]. Also, a step of hole filling was conducted based on Delaunay triangulation.

Once 3D mesh of 3D scans is ready, we establich facial surface extraction and remeshing steps to conduct mesh voxelization. For each obtained volume, we generated 6 views as

TABLE 10. Comparison of the recognition rates with the state-of-the-art on BU3D-FE dataset.

Approach	Accuracy (%)
Soltanpour et al. [21]	95.36
Elaiwat et al. [29]	95.01
Kim et al. [32]	95
Li et al. [33]	96.13
Zihui et al. [78]	96.21
Cai et al. [35]	**98.92**
3D Face ShapeNets (ours)	**97.09**

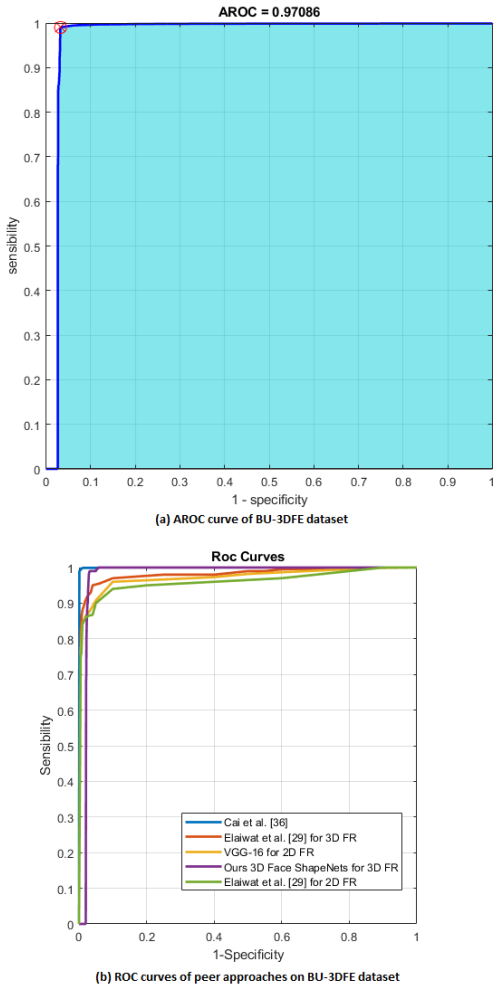


FIGURE 12. ARoc and Roc curves of BU-3DFE dataset: (a) ARoc curve of BU-3DFE, (b) Roc curves of some peer approaches testing 3D FR and 2D FR on BU-3DFE dataset.

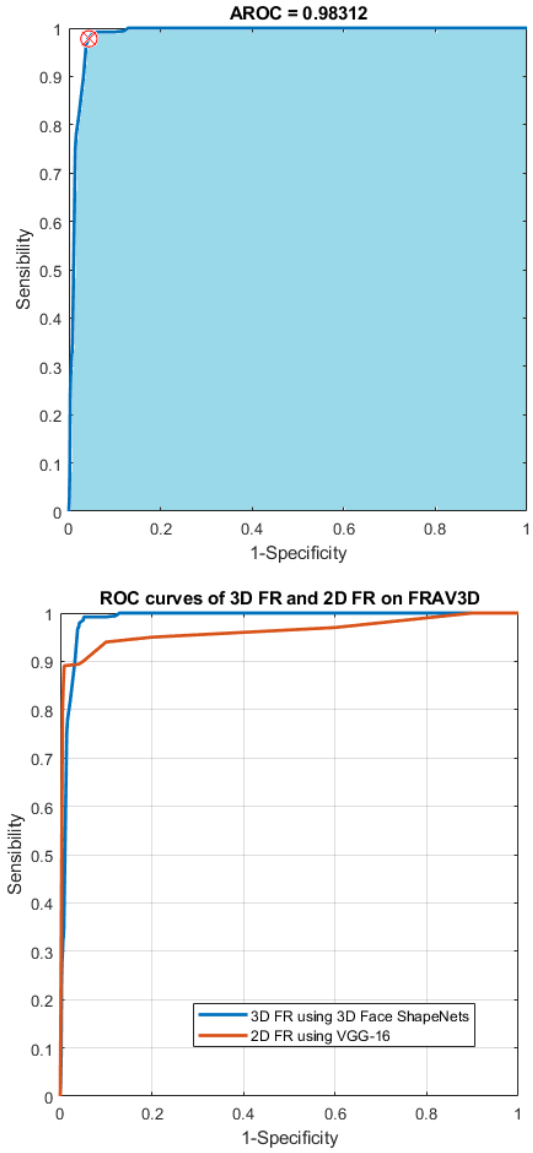


FIGURE 13. ARoc curve of FRAV3D dataset.

TABLE 11. Recognition rates comparison with the state-of-the-art on FRAV3D dataset.

Approach	Accuracy (%)	
Dutta et al. [36]	SpPCANet-1 95.87	SpPCANet-2 96.93
Krishnan et al. [73]	55	
Cheraghian et al. [74]	95	
Bagchi et al. [28]	92.25	
3D Face ShapeNets (ours)	**98.31**	

data augmentation. We randomly used 70% of the scans for training 3D ShapeNet and 30% for testing. The recognition rates are **98.31%** when using 3D ShapeNet (Fig. 13) vs. about **89.15%** with VGG-16. We tested VGG-16 using the 2D images in the dataset. In table 11, a comparison with the state of the art is established.

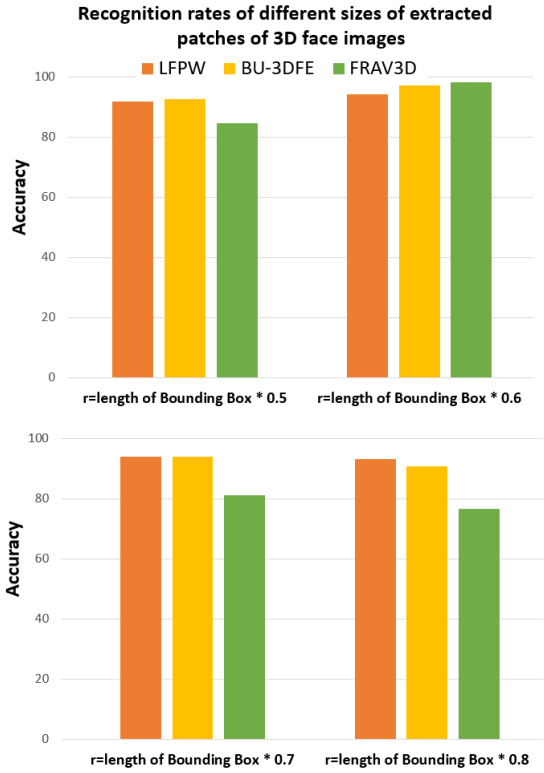


FIGURE 14. Recognition rates of the different sizes of the extracted patches of 3D face images.

D. SELF EVALUATION

Apart from the comparison with the state of the art approaches, we carried out a series of tests to justify our qualitative and quantitative results and the choices of the used parameters and techniques. Besides highlighting the robustness of our contribution through the rates obtained, we would like to emphasize the quality of our work. We start by the chosen radius of the region extracted from the 3D face. We already mentioned that the surface covered by the radius (r = length of Bounding Box * 0.6) is sufficient to extract the salient facial surface, which is suitable for any face shape. We made several tests to determine the extraction radius as shown in Fig. 14.

The choice of the radius of the extracted patch is very important. If r is bigger than the length of Bounding Box multiplied by 0.6, unwanted and unnecessary parts of the face, such as the ears and the neck, will be existing. As a result, the recognition rate can be degraded as shown in Fig.14.

Once we perform salient facial surface extraction of the face, we establish shape correction for faces with missing parts caused by pose variations, especially for 3D faces reconstructed from single 2D image (LFPW dataset). Then, we perform remeshing for mesh regularity. This step is essential to have better voxelization and subsequently a better

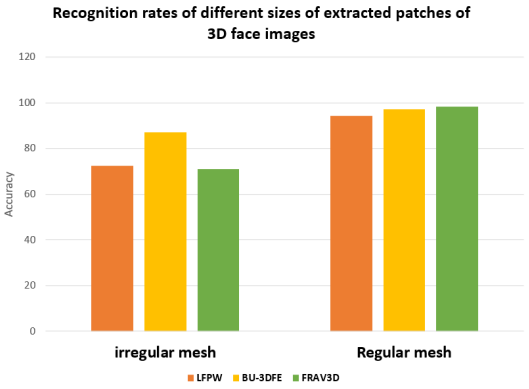


FIGURE 15. Recognition rates when using regular and irregular meshes.

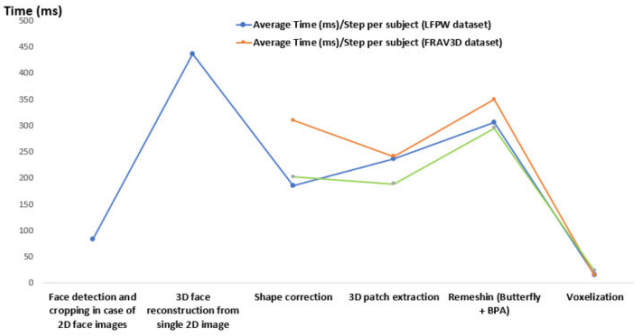


FIGURE 16. Computed time consumed during the preprocessing step per face image.

recognition rate. The histogram in Fig.15 shows that regular mesh increases the recognition rate.

A regular mesh helps to have a perfect voxelization. In fact, each vertex in the space has an occupancy value in the occupancy grid which makes the difference, especially with complex surfaces such as the face in our case.

When comparing the results in BU3D-FE dataset, we notice that the difference between the recognition rate that obtained with regular mesh (97.08%) and that obtained with irregular mesh (87.13%) is less than the difference between the two other datasets. This difference is explained by the fact that meshes of the BU3D-FE dataset are semi-regular, while meshes of FRAV3D dataset need several processing steps for hole filling, denoising, etc.

For the choice of the the number of views for 3D data augmentation, we performed several tests to demonstrate that 12 views are sufficient to give significant information. Indeed, less than 12 views allow no change in the occupancy grid. At each rotation of 15°, the angle of view brings a remarkable change. Indeed, changing the angle of view during yaw axis across 180° is useful, while performing a full 360° turn is not useful because we need views that recover the face surface and not the rear head. Unlike objects, the generated views must cover 360°.

To ensure that our approach is efficient and effective, the time factor is considered. Curves in Fig.16 show the consumed time in each step.

E. DISCUSSION

The experiments carried out proved that the 3D domain gives good results because the geometric information is real and close to reality. Our contribution lies in the use of a 3DCNN that processes 3D volumes.

In this paper, we presented how we adapted 3D ShapeNet to facial recognition. To the best of the authors' knowledge, 3D ShapeNets are used for the first time in face recognition. Indeed, the use of 3D ShapeNet was challenging because voxel grid representation is adequate for objects but not for faces. However, the conducted experiments proved the opposite (Table 9, 10, and 11) which urged us to test other 3D CNNs that work with other 3D representations as mentioned in the state of the art.

Following the experiments we conducted, the recognition rates obtained using 3D ShapeNet are 79.17% (age-related classification on LFPW), 86.68% (gender-related classification on LFPW), 94.25% (LFPW identity identification), 97.09% (on BU-3DFE dataset), and 98.31% (on FRAV3D), respectively. These results are higher than those obtained using 2D CNNs (51.94%, 82.13%, 81.67%, 86% and 89.15%, respectively).

V. CONCLUSION

Our contribution consists in Deep 3D face recognition using 3D ShapeNet. This deep 3D CNN works with 3D volume inputs. We used two types of datasets: a 3D dataset containing images captured by 3D acquisition devices, and a 2D dataset of which we reconstruct the 3D shapes. The first encountered challenge was testing 3D ShapeNet on faces because this 3D CNN was tested only on 3D objects. The second challenge was the use of 2D dataset to reconstruct 3D inputs. The obtained results proved that the reconstruction is considered successful since the recognition rate in 3D domain was higher than the one obtained in 2D domain. It was also proven that working on 3D face recognition is always more effective than 2D face recognition. A future study to test other 3D CNNs and other datasets should be carried out.

ACKNOWLEDGMENT

The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

DECLARATION

Financial Interests: The authors declare they have no financial interests.

Non-Financial Interests: None.

Conflict of Interest: The authors declare that they have no conflict of interest.

REFERENCES

[1] A. Benmohamed, M. Neji, M. Ramdani, A. Wali, and A. M. Alimi, "Feast: Face and emotion analysis system for smart tablets," *Multimedia Tools Appl.*, vol. 74, no. 21, pp. 9297–9322, Nov. 2015.

[2] M. Chakroun, A. Wali, Y. Aribi, and A. M. Alimi, "Video event detection using auto-associative neural network and incremental SVM models," in *Proc. 15th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Dec. 2015, pp. 563–568.

[3] M. Néji, A. Wali, and A. M. Alimi, "The multi-agents architecture for emotion recognition: Case of information retrieval system," *Int. J. Softw. Innov.*, vol. 2, no. 1, pp. 73–85, Jan. 2014.

[4] Z. Kechaou, A. Wali, M. B. Ammar, H. Karray, and A. M. Alimi, "A novel system for video news' sentiment analysis," *J. Syst. Inf. Technol.*, vol. 15, no. 1, pp. 24–44, Mar. 2013.

[5] M. Neji, M. B. Ammar, A. Wali, and A. M. Alimi, "Towards an intelligent information research system based on the human behavior: Recognition of user emotional state," in *Proc. IEEE/ACIS 12th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2013, pp. 371–376.

[6] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DeepFake detection for human face images and videos: A survey," *IEEE Access*, vol. 10, pp. 18757–18775, 2022.

[7] N. A. Al-Humaidan and M. Prince, "Correction to 'a classification of Arab ethnicity based on face image using deep learning approach," *IEEE Access*, vol. 10, p. 32442, 2022.

[8] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh, "A review on face recognition systems: Recent approaches and challenges," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 27891–27922, Oct. 2020.

[9] S. Luo and J. Lu, "GFNet: A gradient information compensation-based face super-resolution network," *IEEE Access*, vol. 10, pp. 8073–8080, 2022.

[10] M. Zhang, R. Liu, D. Deguchi, and H. Murase, "Masked face recognition with mask transfer and self-attention under the COVID-19 pandemic," *IEEE Access*, vol. 10, pp. 20527–20538, 2022.

[11] H. Pranoto, Y. Heryadi, H. L. H. S. Warnars, and W. Budiharto, "Recent generative adversarial approach in face aging and dataset review," *IEEE Access*, vol. 10, pp. 28693–28716, 2022.

[12] F. Albalas, A. Alzu'bi, A. Alguzo, T. Al-Hadhrami, and A. Othman, "Learning discriminant spatial features with deep graph-based convolutions for occluded face detection," *IEEE Access*, vol. 10, pp. 35162–35171, 2022.

[13] D. Wanyonyi and T. Celik, "Open-source face recognition frameworks: A review of the landscape," *IEEE Access*, vol. 10, pp. 50601–50623, 2022.

[14] P. Negri, S. Cumani, and A. Bottino, "Tackling age-invariant face recognition with non-linear PLDA and pairwise SVM," *IEEE Access*, vol. 9, pp. 40649–40664, 2021.

[15] C. Hu, F. Wu, J. Yu, X. Jing, X. Lu, and P. Liu, "Diagonal symmetric pattern-based illumination invariant measure for severe illumination variation face recognition," *IEEE Access*, vol. 8, pp. 63202–63213, 2020.

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

[17] S. Biasotti, E. M. Thompson, L. Barthe, S. Berretti, A. Giachetti, T. Lejemble, N. Mellado, K. Moustakas, I. Manolas, D. Dimou, C. Tortorici, S. Velasco-Forero, N. Werghi, M. Polig, G. Sorrentino, and S. Hermon, "SHREC'18 track: Recognition of geometric patterns over 3D models," in *Proc. Eurograph. Workshop 3D Object Retr.* vol. 2, 2018, pp. 1–8.

[18] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger, "Differentiable volumetric rendering: Learning implicit 3D representations without 3D supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3504–3515.

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

[20] X. Deng, F. Da, and H. Shao, "Efficient 3D face recognition using local covariance descriptor and Riemannian kernel sparse coding," *Comput. Electr. Eng.*, vol. 62, pp. 81–91, Aug. 2017.

[21] S. Soltanpour and Q. M. J. Wu, "Weighted extreme sparse classifier and local derivative pattern for 3D face recognition," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3020–3033, Jun. 2019.

[22] L. Shi, X. Wang, and Y. Shen, "Research on 3D face recognition method based on LBP and SVM," *Optik*, vol. 220, Oct. 2020, Art. no. 165157.

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

[24] Y. Tang, X. Sun, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "3D face recognition with asymptotic cones based principal curvatures," in *Proc. Int. Conf. Biometrics (ICB)*, May 2015, pp. 466–472.

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

[26] J. B. C. Neto and A. N. Marana, "Utilizing deep learning and 3DLBP for 3D face recognition," in *Proc. Iberoamerican Congr. Pattern Recognit.* Cham, Switzerland: Springer, 2017, pp. 135–142.

[27] M. Peter, J.-L. Minoi, and I. H. M. Hipiny, "3D face recognition using kernel-based PCA approach," *Computational Science and Technology*. Singapore: Springer, 2019, pp. 77–86.

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

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

[30] Y. Liang, Y. Zhang, and X.-X. Zeng, "Pose-invariant 3D face recognition using half face," *Signal Process., Image Commun.*, vol. 57, pp. 84–90, Sep. 2017.

[31] K. Dutta, D. Bhattacharjee, M. Nasipuri, and O. Krejcar, "Complement component face space for 3D face recognition from range images," *Appl. Intell.*, vol. 51, no. 4, pp. 2500–2517, Apr. 2021.

[32] D. Kim, M. Hernandez, J. Choi, and G. Medioni, "Deep 3D face identification," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 133–142.

[33] H. Li, J. Sun, and L. Chen, "Location-sensitive sparse representation of deep normal patterns for expression-robust 3D face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 234–242.

[34] J. Luo, F. Hu, and R. Wang, "3D face recognition based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2019, pp. 1576–1581.

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

[36] K. Dutta, D. Bhattacharjee, and M. Nasipuri, "SpPCANet: A simple deep learning-based feature extraction approach for 3D face recognition," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31329–31352, Nov. 2020.

[37] Z. You, T. Yang, and M. Jin, "Multi-channel deep 3D face recognition," 2020, *arXiv:2009.14743*.

[38] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009.

[39] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "FaceWarehouse: A 3D facial expression database for visual computing," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 3, pp. 413–425, Mar. 2014.

[40] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, and X. Tong, "O-CNN: Octree-based convolutional neural networks for 3D shape analysis," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, Aug. 2017.

[41] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1912–1920.

[42] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5115–5124.

[43] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[44] H. Maron, M. Galun, N. Aigerman, M. Trope, N. Dym, E. Yumer, V. G. Kim, and Y. Lipman, "Convolutional neural networks on surfaces via seamless toric covers," *ACM Trans. Graph.*, vol. 36, no. 4, p. 71, 2017.

[45] A. Garcia-Garcia, F. Gomez-Donoso, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla, and J. Azorin-Lopez, "PointNet: A 3D convolutional neural network for real-time object class recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1578–1584.

[46] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1–4.

[47] S. P. T. Reddy, S. T. Karri, S. R. Dubey, and S. Mukherjee, "Spontaneous facial micro-expression recognition using 3D spatiotemporal convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[48] J. Haddad, O. Lézoray, and P. Hamel, "3D-CNN for facial emotion recognition in videos," in *Proc. Int. Symp. Vis. Comput.* Cham, Switzerland: Springer, 2020, pp. 298–309.

[49] B. Hasani and M. H. Mahoor, "Facial expression recognition using enhanced deep 3D convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 30–40.

[50] W. Xie, L. Liang, Y. Lu, H. Luo, and X. Liu, "Deep 3D-CNN for depression diagnosis with facial video recording of self-rating depression scale questionnaire," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 2007–2010.

[51] M. Jabberi, A. Wali, B. B. Chaudhuri, and A. M. Alimi, "68 landmarks are efficient for 3D face alignment: What about more?" *Multimedia Tools Appl.*, pp. 1–35, Apr. 2023.

[52] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2879–2886.

[53] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Mille, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces Real-Life Images, Detection, Alignment, Recognit.*, 2008, pp. 1–14.

[54] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*.

[55] M. Kostinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2144–2151.

[56] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2930–2940, Dec. 2013.

[57] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.

[58] *Face Recognition and Artificial Vision Group FRAV2D Face Database*. Accessed: Apr. 23, 2022. [Online]. Available: http://www.frav.es

[59] A. Moreno and A. Sánchez, "GavabDB: A 3D face database," Presented at the 2nd COST275 Workshop Biometrics Internet, Vigo, Spain, 2004.

[60] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D face recognition database," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, May 2010, pp. 97–100.

[61] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Proc. Eur. Workshop Biometrics Identity Manag.* Berlin, Germany: Springer, 2008, pp. 47–56.

[62] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 211–216.

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

[64] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 947–954.

[65] C. Conde, A. Serrano, and E. Cabello, "Multimodal 2D, 2.5D & 3D face verification," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 2061–2064.

[66] C. Mandal, H. Qin, and B. C. Vemuri, "Dynamic modeling of butterfly subdivision surfaces," *IEEE Trans. Vis. Comput. Graphics*, vol. 6, no. 3, pp. 265–287, Jul./Sep. 2000.

[67] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE Trans. Vis. Comput. Graphics*, vol. 5, no. 4, pp. 349–359, Oct. 1999.

[68] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, May 2006.

[69] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Sep. 2002.

[70] T. Tieleman and G. Hinton, "Using fast weights to improve persistent contrastive divergence," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 1033–1040.

[71] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[72] J. Vollmer, R. Mencl, and H. Mueller, "Improved Laplacian smoothing of noisy surface meshes," *Comput. Graph. Forum*, vol. 18, no. 3, pp. 131–138, 1999.

[73] P. Krishnan and S. Naveen, "RGB-D face recognition system verification using Kinect and FRAV3D databases," *Proc. Comput. Sci.*, vol. 46, pp. 1653–1660, Jan. 2015.

[74] A. Cheraghian, K. Faez, H. Dastmalchi, and F. B. Oskuie, "An efficient multimodal face recognition method robust to pose variation," in *Proc. IEEE Symp. Comput. Informat.*, Mar. 2011, pp. 431–435.

[75] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on CIFAR-10," *Unpublished Manuscript*, vol. 40, no. 7, pp. 1–9, 2010.

[76] J. Zhang, Y. Jiang, X. Hu, and J. Shen, "Convolutional restricted Boltzmann machine based on fast persistent contrastive divergence," *Comput. Eng.*, vol. 9, 2016.

[77] M. A. Carreira-Perpinan and G. Hinton, "On contrastive divergence learning," in *Proc. Int. Workshop Artif. Intell. Statist.*, 2005, pp. 33–40.

[78] Z. Zhang, C. Yu, S. Xu, and H. Li, "Learning flexibly distributional representation for low-quality 3D face recognition," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 4, pp. 3465–3473.

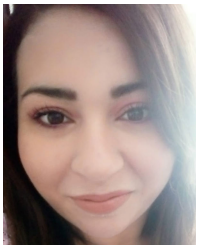
[79] M. A. Talab, N. A. Qahraman, M. M. Aftan, A. H. Mohammed, andM. D. Ansari, "Local feature methods based facial recognition," in *Proc. Int. Congr. Hum.-Comput. Interact., Optim. Robot. Appl. (HORA)*, Jun. 2022, pp. 1–5.

[80] A. S. Gezawa, Y. Zhang, Q. Wang, and L. Yunqi, "A review on deep learning approaches for 3D data representations in retrieval and classifications," *IEEE Access*, vol. 8, pp. 57566–57593, 2020.

[81] R. Poojary and A. Pai, "Comparative study of model optimization techniques in fine-tuned CNN models," in *Proc. Int. Conf. Electr. Comput. Technol. Appl. (ICECTA)*, Nov. 2019.

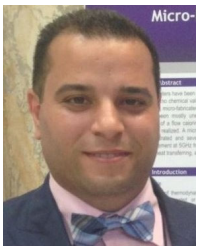
[82] S. Tammina, "Transfer learning using VGG-16 with deep convolutional neural network for classifying images," *Int. J. Sci. Res. Publications*, vol. 9, no. 10, pp. 143–150, 2019.

[83] M. Jabberi, A. Wali, and A. M. Alimi, "Generative data augmentation applied to face recognition," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2023, pp. 242–247.



MARWA JABBERI (Member, IEEE) was born in Bizerte, Tunisia, in 1990. She received the master's degree in automatic reasoning systems from the Faculty of Sciences of Monastir, Tunisia, in 2014. She is currently pursuing the Ph.D. degree in computer science with the REsearch Groups on Intelligent Machines (REGIM Laboratory), ENIS. Her research focuses on applying intelligent methods (deep neural networks, feature extraction, and recognition algorithms) to deep machine learning, computer vision, face image processing, intelligent pattern recognition, and analysis of large scale complex systems.

ALI WALI (Senior Member, IEEE) received the Ph.D. degree in engineering computer systems from the National School of Engineers of Sfax, in 2013. He is an Associate Professor of signal and image processing with ISIM, University of Sfax. He is also a member of the REsearch Groups on Intelligent Machines (REGIM Laboratory). His research interests include computer vision and image and video analysis. These research activities are centered around video events detection and pattern recognition. He was a member of the Organizing Committee of the International Conference on Machine Intelligence ACIDCA-ICMI2005, the Third IEEE International Conference on Next Generation Networks and Services (NGNS2011), the Fourth International Conference on Logistics (LOGISTIQUA2011), the International Conference on Advanced Logistics and Transport (ICALT'2013, ICALT'2014, and ICALT'2015), the 13th International Conference on Document Analysis and Recognition (ICDAR 2015), the International Conference on Hybrid Intelligent Systems (HIS'2013), and the International Conference on Information Assurance and Security (IAS'2013).



BILEL NEJI was born in Tunisia, in May 1983. He received the B.Eng. degree in electrical engineering from the National Engineering School of Sfax, Tunisia, in cooperation with Valenciennes University, France, in 2007, the M.S. degree in new technologies in computer systems from the National Engineering School of Sfax, in cooperation with Lille University, France, in 2008, the first Ph.D. degree in electrical engineering, focused on scientific satellites' subsystems from the National Engineering School of Sfax, in cooperation with Wurzburg University, Germany, in 2014, and the second Ph.D. degree in electrical engineering, focused on MEMS and micro/nano sensors design and fabrication from the State University of New York at Buffalo, USA, in 2015. He has Co-Founded BAK USA Technologies Corporation, NY, USA, in 2014, where he served as the Director of engineering until 2018. He joined the American University of the Middle East (AUM), Kuwait, in September 2018, as an Assistant Professor. He is currently an Associate Professor with AUM, where he also serves as the Research Office Coordinator of the College of Engineering and Technology. He has been conducting research in different areas, including embedded systems, MEMS and sensors design, artificial intelligence, and renewable energy. He received the Fulbright Science and Technology Award from the U.S. Department of States, USA, in 2011, and several other prestigious awards worldwide.

TAHA BEYROUTHY received the Ph.D. degree in micro and nano electronics from the Grenoble Institute of Technology, in 2009, and the degree in engineering education from IMT-Atlantique (Télécom-Bretagne). He joined the American University of the Middle East (AUM), Kuwait, in November 2013, as an Assistant Professor, and was promoted to an Associate Professor, in 2017. He has been the Dean of engineering and technology with AUM, since September 2017. He has been instrumental in AUM growth of higher education through his broad experience in academic leadership and commitment to both a student centered education and a technologically empowered teaching and learning environment. As an Associate Professor of electrical engineering, he has authored/coauthored more than 100 peer-reviewed publications in micro and nano electronics, robotics, artificial intelligence, and applied physics.

ADEL M. ALIMI (Senior Member, IEEE) was born in Sfax, Tunisia, in 1966. He received the degree in electrical engineering, in 1990, and the Ph.D. and H.D.R. degrees in electrical and computer engineering, in 1995 and 2000, respectively. He is a Professor of electrical and computer engineering with the University of Sfax. His research interests include the applications of intelligent methods (neural networks, fuzzy logic, and evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. His research focuses on intelligent pattern recognition, learning, analysis, and intelligent control of large scale complex systems. He is the Founder and the Chair of many IEEE Chapter in Tunisia Section. He is the IEEE Sfax Subsection Chair, in 2011, the IEEE ENIS Student Branch Counselor, in 2011, the IEEE Systems, Man, and Cybernetics Society Tunisia Chapter Chair, in 2011, and the IEEE Computer Society Tunisia Chapter Chair, in 2011. He is also an Expert Evaluator of the European Agency for Research. He was the General Chairman of the International Conference on Machine Intelligence ACIDCA-ICMI'2005 and ACIDCA-ICMI'2000. He is an Associate Editor and a member of the editorial board of many international scientific journals, such as IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Neurocomputing*, *Neural Processing Letters*, *International Journal of Image and Graphics*, *Neural Computing and Applications*, *International Journal of Robotics and Automation*, and *International Journal of Systems Science*. He was a Guest Editor of several special issues of international journals, such as *Fuzzy Sets and Systems*, *Soft Computing*, *Journal of Decision Systems*, *Integrated Computer Aided Engineering*, and *Systems Analysis, Modeling, and Simulations*.