Multimedia Tools and Applications (2020) 79:26517–26547
https://doi.org/10.1007/s11042-020-09331-5

Voxel-based 3D occlusion-invariant face recognition using game theory and simulated annealing

Sahil Sharma[1] · Vijay Kumar[2]

Received: 8 December 2019 / Revised: 13 June 2020 / Accepted: 9 July 2020 /
Published online: 16 July 2020

Abstract

A novel voxel-based occlusion-invariant 3D face recognition framework (V3DOFR) based on game theory and simulated annealing is proposed. In V3DOFR approach, 3D meshes are converted to voxel form of sizes 4^3, 8^3, and 16^3. After that, locality preserving projection-based embeddings are computed for removing the sparseness of voxels and generating consistent linear embedding per mesh with size 64×3, 128×3, and 256×3, respectively. The generator of triplets provides the triplets of sizes 64x3x3, 128x3x3, and 256x3x3. The simulated annealing is used to check the threshold value of adversarial triplet loss generated after ensembling losses of different grid sizes. The proposed framework is compared with four well-known methods using three face datasets, namely, Bosphorus, UMBDB, and KinectFaceDB. The performance evaluation has been done using four different cases of experimentations, viz. voxel based face recognition, occlusion invariant face recognition, landmarks based 3D face recognition, and 3D mesh based face recognition. Seven evaluation metrics are used to compare the proposed technique with other methods. The proposed method provides better accuracy and computation time over the other existing techniques in the majority of cases.

Keywords 3D mesh · Voxelization · Adversarial triplet loss · Generator · Discriminator · Simulated annealing

✉ Sahil Sharma
sahil301290@gmail.com

Vijay Kumar
vijaykumarchahar@gmail.com

[1] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India

[2] Computer Science and Engineering Department, National Institute of Technology, Hamirpur, India

1 Introduction

3D face recognition is widely used throughout the world due to the availability of easily collectable 3D data and capabilities of computation with the availability of highly economical graphical processing units (GPUs). However, acquiring 3D images are harder as compared to 2D scans. Therefore, the number of images is limited in public databases [25, 86, 90]. In [90], a high resolution spontaneous 3D dynamic facial expression database is presented. This work supports 3D spatiotemporal features exploration in subtle face expression. In [86], high resolution data acquisition is done using 3D dynamic imaging system setup. There are total of 101 number of subjects, six unique expressions, 606 number of 2D texture videos, 606 number of 3D model sequences, and approximately 60,600 3D models. In [25], 3D face recognition is improved using multi-instance enrollment representation. The experiments were performed on ND-2006 3D face dataset [57], that consists of 13,450 3D images. There are various techniques available in the literature for handling 3D mesh data, RGB-D image, or point cloud data [23, 64, 99]. ElSayed et al. [23] presented a robust method for detection of skin using 3D colored point clouds. This method is extended to solve 3D face detection problem by building a weighted graph for the initial 3D colored point clouds. A linear programming algorithm is used for predicting model using data mining approach and classifying the graph regions of skin versus non-skin regions. Zhou et al. [99] presented a dense 3D face decoding method using a non-linear 3D morphable model (3DMM) by training over joint texture and shape autoencoders using direct mesh convolutions. It is shown in [99] that how these autoencoders are usable in training very light weight models performing Coloured Mesh Decoding (CMD) at speed of over 2500 FPS. Pham et al. [64] presented a novel robust hybrid 3D face tracking framework from RGBD videos. It is capable of tracking head pose and face actions without any intervention or recalibration from a user.

Some well-known methods for handling 3D features are slower in computation time as compared to deep learning techniques [48, 50, 76]. Spreeuwers [76] presented a 3D face registration technique based on intrinsic coordinate system of the face. Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) is used for feature extraction along with matching score of likelihood ratio. The overall method takes 2.5 s per image, which is too slow as compared to the technique proposed in this paper. Li et al. [50] presented 3D face recognition technique by extending SIFT-like matching framework to mesh data. Lei et al. [48] represented the facial scan using keypoint-based multiple triangle statistics (KMTS), which is a robust method to partial facial data, pose variations, and facial expressions. An approach called two-phase weighted collaborative representation classification (TPWCRC) is used. Experiments were performed on Bosphorus [71], UMBDB [17], GavabDB [55], SHREC 2008 [82], BU-3DFE [85], and FRGC v2.0 [66] datasets. There are various challenges in 3D face recognition viz. pose, occlusion, expression, lighting, etc. These variations affect intra-class recognition capabilities of 3D face recognition system [41].

Based on the above mentioned issues, a voxel-based 3D face recognition system is proposed that utilize the basic concepts of locality preserving projections (LPP), triplet loss, simulated annealing, and game theory. The reason behind the use of LPP is to remove the sparseness of meshes with a non-uniform number of voxels. LPP is chosen over PCA due to representing high dimensional data in low dimension, LPP is computed by optimal linear approximation to eigenfunctions of Laplace Beltrami operator on the manifold whereas PCA projects the data along with maximal variance directions [33]. The triplet loss training reduces the distance between the intraclass faces and maximizes the distance between different class faces. It helps in increasing the reliability of the system in face identification. Simulated annealing is used for minimizing the error rate by using the probability-based random threshold value. Generator and discriminator are part of game theory that helps in the correct selection of triplet generated based on simulated annealing. The combined effect of these techniques makes the proposed method robust towards the occlusion in 3D face recognition.

The main contributions of this paper are as follows.

1. The proposed approach utilizes generator and discriminator for voxel-based face recognition.
2. A deep learning and simulated annealing based framework is proposed for voxel-based 3D occlusion invariant face recognition (V3DOFR).
3. The proposed approach is validated using three standard datasets with a significant amount of pose and occlusion variation.
4. The proposed technique is compared with other state-of-the-art methods for voxel-based face recognition, occlusion invariant face recognition, 3D landmarks based face recognition, and 3D mesh-based face recognition.

The remaining structure of this paper is organized as follows. Section 2 presents the background work done in the field of face recognition. The proposed research framework is presented in Section 3. Section 4 discusses the experimental results in detail. Section 5 presents the future work. The concluding remarks are drawn in Section 6.

2 Background

In this section, the basic concepts of deep learning-based face recognition, voxelization, locality preserving projections (LPP), triplet loss, game theory, and simulated annealing are discussed. Thereafter, the related work is discussed.

2.1 Preliminary

2.1.1 Deep learning-based 3D face recognition

Training and testing are the two main phases of deep learning-based 3D face recognition (see Fig. 1). There are two sub-phases in training phase namely, pre-processing and deep learning. During pre-processing, 3D face acquisition is made by three methods such as RGB-D depth image, 3D face mesh image, and 3D point cloud image [61]. Once 3D face is acquired, face alignment as well as registration is done for maximum utilization of the available information. There are three methods of proceeding with 3D face after alignment. First, coarse-detail based facial landmark detection is one of the fastest method. Facial landmarks lack the finer details of the face. To overcome this, voxelization is used that includes fine-details of a 3D face. Voxelization is a slower process in contrast to landmarks detection. The third method is to use a 3D object as a whole, in the form of 3D RGB-D image, mesh image, or the point cloud image. After the completion of pre-processing phase, there is an availability of deep learning models. There are different types of models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders (AEs), generative adversarial networks (GANs), and reinforcement learning (RL) [2, 3]. CNN's and RNNs are used in supervised learning for images and text, respectively. AEs and GANs are used in semi-supervised learning. RL is used in unsupervised learning.

There are two sub-phases in testing phase namely, validation and verification. In validation, the testing dataset is processed through alignment. Face registration and trained deep learning model is used to predict the array of classes corresponding to all images in testing dataset. The accuracy of face recognition model is computed as follows [37].

$$\text{Accuracy} = \frac{\textit{Number}\ \text{of images correctly predicted}}{\textit{Total}\ \text{number of images}}\ \text{x}\ 100 \tag{1}$$

In verification, the query image is processed through face alignment and registration. The trained deep learning model is used to predict the class of image. Finally, a similarity score is calculated in comparison to the predicted class images [14].

$$\text{Similarity Score} = \sum_{i=1}^{N} \frac{corr(Query\ \text{Image}, \mathrm{I}_i)}{N} \tag{2}$$

2.1.2 Voxelization

Voxel representation is widely used in multiple fields viz. computer graphics, computational science, real-time computer vision, and 3D shape matching. Dynamic modeling requires voxelization or real-time scan conversion. In this process, the triangular mesh is used to create voxel representation from the input surface [60].

Let point O be arbitrary origin point, and G be a polyhedron with triangular faces t_1, t_2, t_3, …, t_n, then $H = \{H_1, H_2, H_3, \ldots, H_n\}$ be covering of G with 3D tetrahedra. H_i is defined by O and triangular faces t_i [58]. Point A is considered to be inside the polyhedron G iff

$$\textstyle\sum_i \mathrm{sign}(\mathrm{H}_i)\ \mathrm{incl}(\mathrm{H}_i, A) > 0 \tag{3}$$

where sign($\mathrm{H_i}$) is true when $\mathrm{H_i} > 0$, and incl($\mathrm{H_i}$, A) is true for all $\mathrm{A} \in \mathrm{H_i}$.

2.1.3 Locality preserving projections

Suppose there are large n-dimensional vectors of data points. The intrinsic property of data is used for dimensionality reduction of these large vectors. Locality preserving projection (LPP) builds a graph that consists of the neighborhood information of dataset [33] and solves the

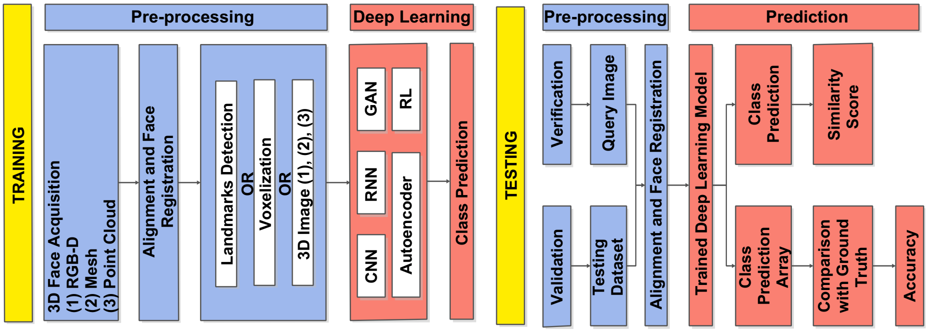

Fig. 1 A general framework of deep learning-based 3D face recognition

linear dimensionality reduction problem. LPP is a linear approximation of non-linear Laplace Eigenmap and is as follows [9, 33].

```
Algorithm: Locality Preserving Projections
Input: Graph G with P Nodes
Output: l-dimensional vector
    1. Nodes a and b are connected iff
        a) Two nodes are close according to Euclidean norm $R^n$.
           $|| x_a - x_b ||^2 < \in$        //$\in$ is number of neighbours
        b) Nodes a and b are among the k-nearest neighbors of each other.
    2. Weighting of edges
        a) In case of given parameter p, p $\in$ R, if nodes a and b are connected,
           $W_{ab} = e^{-\frac{||x_a - x_b||^2}{p}}$
        b) In case of a and b are connected by only one edge,
           $W_{ab} = 1$
    3. Computation of eigenvalues and eigenvectors for generalized eigenvector problem
        a) $ZL_m Z^T a = \lambda ZMZ^T a$
           Where M is the diagonal matrix with column sums of W. $L_m$ is the Laplacian matrix,
           $L_M = M - W$. $Z_i$ is the $i_{th}$ column of matrix Z.
        b) Suppose $d_1, d_2, ...., d_{l-1}$ be the column vector solution of equation 3(a), ordered
           corresponding to eigenvalues $\lambda_0, \lambda_1, ...., \lambda_{\lambda-1}$, final embedding is as follows.
           $x_i \rightarrow y_i = S^T z_i$, $S = (S_0, S_1, ...., S_{l-1})$
           where $y_i$ is l-dimensional vector, and S is n x l matrix.
```

In locality preserving projections algorithm, Graph G with P nodes are connected according to the Euclidean norm and k-nearest neighbors of two nodes (see Step 1). In Step 2, weights are assigned to nodes. In the final step, the computation of final *l*-dimensional vector is calculated based on the generalized eigenvector problem.

2.1.4 Triplet loss

Triplet loss uses face embeddings as vectors. It chooses three embeddings namely, Anchor (A), Positive (P), and Negative (N) from the dataset such that A and P belong to the same class, and N belongs to a different class. A, P, and N are selected randomly based on three categories viz. easy triplets, hard triplets, and semi-hard triplets. Easy triplets (see Eq. 4) has a loss of 0. In hard triplets (see Eq. 5), negative embedding is closer to anchor embedding as compared to positive embedding. In semi-hard triplets (see Eq. 6), the negative embedding is not closer than the positive embedding but still has positive loss [73].

$$d(A,P) + margin < d(A,N) \tag{4}$$

$$d(A,N) < d(A,P) \tag{5}$$

$$d(A,P) < d(A,N) < d(A,P) + \mathrm{margin} \tag{6}$$

The loss of a triplet (A, P, N) is defined as [73].

$$L = \max(d(A, P) - d(A, N) + margin, 0) \tag{7}$$

The main objective of triplet loss is to minimize the loss by pushing d(A, P) → 0 and d(A, N) > d(A, P) + margin by triplet loss training. Figure 2 represents the concept of triplet loss using three images given as input in form of A, P, and N to deep learning model for triplet embeddings and triplet loss training.

2.1.5 Game theory

Game theory is a term used jointly with generative adversarial networks (GANs). GANs [29] are one of the types of generative models. Let $P_{data}(I)$ be the distribution of a real image I and $P_J(J)$ be the distribution of the input. Let generator G(z) capture the P_{data} distribution by using an adversarial process. The discriminator D distinguishes between real images and generated images. The formulation of the adversarial process in the form of a minimax game (see Eq. 8).

$$\min_G \max_D E_{I \sim P_{data}} \left[\log D(I) + \mathrm{E}_{Z \sim P_Z}[\log(1 - D(G(z)))]\right] \tag{8}$$

Theoretically, the global optimum $\mathrm{P}_{\mathrm{G(Z)}} = \mathrm{P}_{\mathrm{data}}$ [30] is reached by the minimax game on reaching Nash equilibrium [69] by the adversarial process. Recently, AttGAN [34] has achieved facial attribute editing as well as gender and age transformation (see Fig. 3(a) and (b)) viz. reconstruction of blond hair, eye glass, changing expression, makeup, etc. Face aging with conditional GANs [4] achieved remarkable results in generating faces of different age from a single image (see Fig. 3(c)).

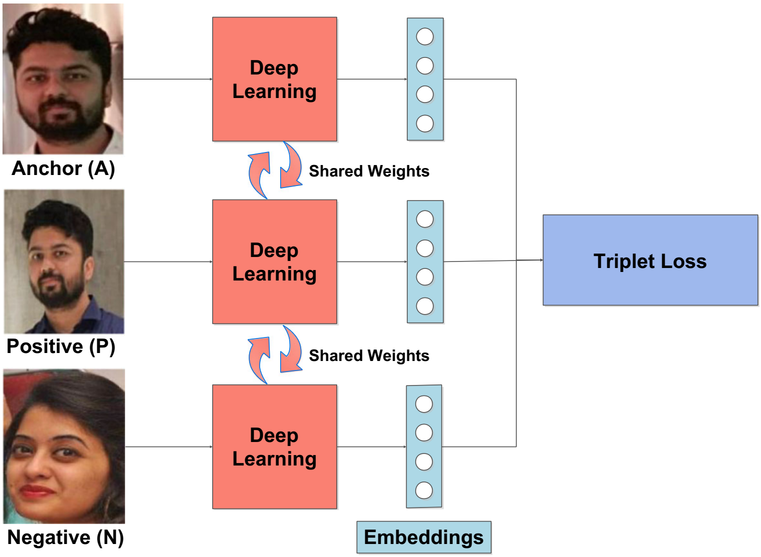

Fig. 2 Concept of triplet loss technique

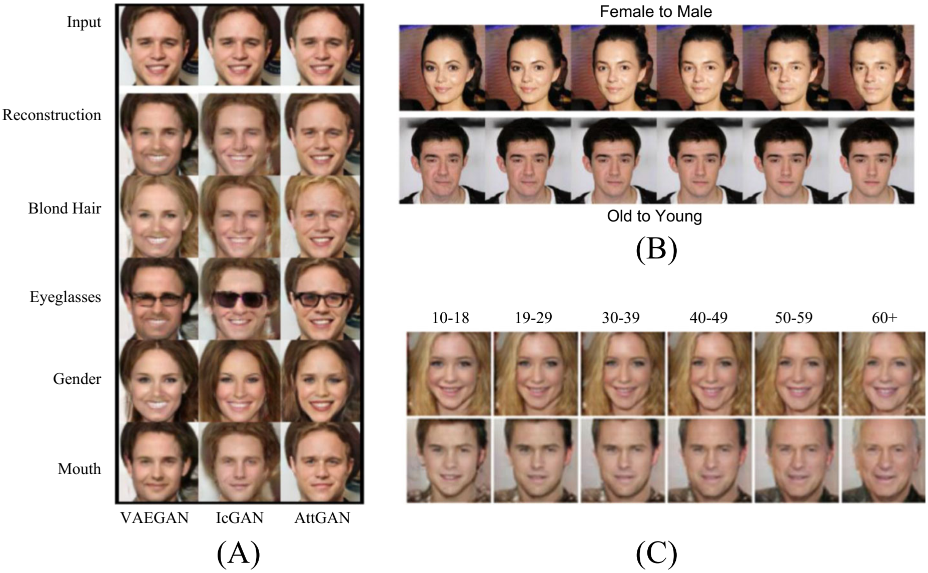

Fig. 3 **a** Comparison of transforming attributes using VAE-GAN [46], IcGAN [63], and AttGAN [34], **b** Gender and age transformation using AttGAN, and **c** Transformation of face to any age using Age Conditional GAN [4]

2.1.6 Simulated annealing

Simulated annealing (SA) optimization algorithm is based on metallurgical practices in which a particular material is heated at high temperature, and then it is brought to low temperature gradually. Shifting of atoms become unpredictable when heating a material at high temperature. It helps in the elimination of impurities as the material takes pure crystal form after cooling. In terms of optimization, SA introduces a degree of randomness, which may take the solution from better to worse in an attempt to escape local minima and increasing the probability of achieving global optimum [44]. The applications of SA are diverse [7, 13, 52] by single criterion optimization [8].

Figure 4 holds four states A, B, C, and D having different energy. The main target is to find a path having maximum energy by using simulated annealing algorithm to traverse every state exactly once. For understanding purposes, the four states have been connected in two ways, clockwise and anticlockwise. The total sum of energies by clockwise traversing of all states is 35, whereas the total sum of energies by anticlockwise traversing is 70. Hence, the maximum energy state path is selected by anticlockwise traversing.

```
Algorithm: Simulated Annealing
Input: Random Initial state
Output: Path covering all states with maximum energy
1.  $S = S_{init}$                          //Selection of initial state randomly
2.  For $T = T_{max}$ to $T_{min}$ do       //New temperature in every iteration
        a. $E_S = E(S)$                     //Current state energy
        b. $N = next(S)$                    //New state
        c. $E_N = E(N)$                     //New state energy
        d. $\Delta E = E_N - E_S$           //Change in energy
        e. If($\Delta E > \Theta$)          //Maximization problem
        f.     $S = N$                      //New state selected
        g. Else if ($e^{\Delta E/T} > rand(0,1)$)
        h.     $S = N$                      //New state with probability $e^{\Delta E/T}$
        i. End if
    End for
```

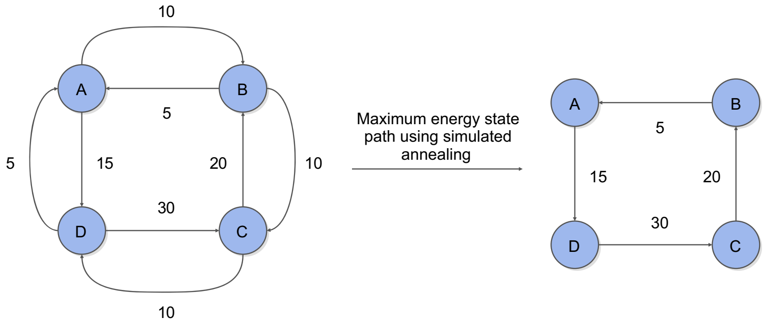

Fig. 4 State energy maximization using simulated annealing

In a simulated annealing algorithm, the initial state is chosen randomly (see Step 1). E_s acts as current state energy. The new state becomes the current state if there is a positive energy change, and otherwise, the new state is chosen with probability $e^{\Delta E/T}$ (see Step 2). During the process, temperature T is decreased gradually so that solution converges towards the global optimum.

2.2 Related work

Kim et al. [42] proposed a novel 3D face recognition algorithm using a face expression augmentation technique alongwith deep convolutional neural network. They used 2.5D or depth images as 3D face images and transfer learning on FRGC [66], CASIA 3D [12], BU-3DFE [85], Bosphorus [71], and 3D-TEC [78] datasets. They presented a technique of augmenting facial expressions from single 3D face image. VGG-16 model has been used to implement the transfer learning. They claimed the rank-1 accuracy for face recognition is 99.2% for Bosphorus. However, it would fail on voxel based, mesh based, or landmark based 3D face recognition technique when the input data is in sequential nature.

Gilani et al. [101] proposed a technique to generate millions of 3D facial images of unique identities by simultaneously interpolating between the facial identity and facial expression spaces to close the gap between the sizes of 2D and 3D datasets. There may be a loss in the depth factor due to the conversion of 3D into 2D images but the augmentation makes up for the loss. In closed and open world recognition scenarios, the proposed FR3DNet outperforms the existing face recognition algorithms. The main advantage of [101] is that it helps in building bigger 3D datasets as compared to the standard 3D datasets. This work is done on 3D images generated from 2D images. Similar to [42], this method fails on voxel based face recognition.

Korshunov and Marcel [45] proposed a public dataset, namely, Deepfake, generated with VidTIMIT database [70]. The main objective of this dataset is to help in generating the swapped faces of two people from videos using generative adversarial networks (GANs). This work is done on 2D face videos. They emphasized that the quality of video is significantly impacted by training and blending parameters. It is observed that the VGG-Net and FaceNet are in jeopardy due to Deepfake videos. The error rate obtained from the FaceNet was 8.97%. The Deepfakes generated from GANs are challenging for both the face detection and recognition systems. The face swap technology based on GANs provided greater challenge to 2D face recognition. This work if presented on voxel face videos would be a challenging for 3D face recognition.

Gecer et al. [28] came up with a novel 3D Morphable Models (3DMM) fitting strategy, which is based on generative adversarial networks (GANs) and a differentiable renderer. The novel cost function integrated various content losses on deep identity features from a face recognition network. The high fidelity 3D Face Reconstruction was achieved by using non-linear energy based cost optimization, GAN texture model, differentiable renderer, cost function, and model fitting techniques. During the fitting process, Adam solver was used for optimization. Abrevaya et al. [1] presents a GAN based 3D face modeling novel architecture which combines a 3D generator with a 2D discriminator leveraging the conventional CNNs. The feature loss, identity loss, and expression loss are calculated by the discriminator to give the real or fake output. Four publicly available 3D face datasets have been used namely BU-3DFE [85], Bosphorus [71], BP4D-Spontaneous [90], and BP-4DFE [89].

Patil et al. [62] presents a survey on 3D face recognition. It provided an extensive review on 3D face recognition in terms of feature detection, the classifiers, 3D face databases, types of 3D facial data acquisition techniques viz. stereo acquisition, laser beam scanning, and fringe pattern acquisition using structured light. Different 3D face representations, namely, point cloud representation, 3D mesh representation, and depth image representation are discussed. Different registration techniques of 3D faces as iterative closest point (ICP) algorithm, spin images, simulated annealing, and intrinsic coordinate system are discussed with their pros and cons. Wu et al. [81] extracted the features from the whole 3D model. For 3D object shapes, a volumetric representation is used. Based on the mesh surface, each voxel location contains a binary value of 0 or 1 with a grid size of 30^3. Voxel grids hold vast information in terms of facial density and texture. This method is better than the depth image technique. Moreover, voxels can be directly used in training 3D convolutional neural networks (3DCNNs) and 3D generative adversarial networks (3DGANs).

Rathgeb et al. [68] presented an overview of impact and detection of facial beautification in face recognition. The plastic surgery, facial retouching, facial cosmetics are common these days. Due to these beautifications, the face recognition based biometrics become an enormous challenge. Facial recognition is used in mobile phones unlocking, payment applications, automated border control etc. The challenges were presented in this work. All the work discussed in this paper is 2D in nature and lacks discussion on 3D face recognition and challenges. Hassaballah and Aly [31] discussed about the significant challenges, which are faced while building a face recognition system for the real world. 3D face recognition and video-based face recognition have been discussed in the work, however, deep learning based techniques are not mentioned at all.

Scherhag et al. [72] presented the survey on face recognition systems under morphing attacks. The generalizability of deep face recognition systems have increased the vulnerability against attacks. The morphing on 2D faces using correspondence, warping, and blending are discussed. This work does not discuss morphing over 3D face images. Ding and Tao [19] discussed 2D face image based pose-invariant face recognition (PIFR). PIFR methods are grouped in four categories viz. pose-robust feature extraction approaches, face synthesis approaches, multi-view subspace learning approaches, and hybrid approaches. The main challenge of face recognition under different poses is self-occlusion because of non-frontal pose. Other challenges are resolution of an image, illumination in an image, and expression.

Bowyer et al. [11] presented a survey of 3D and multi-modal 3D + 2D face recognition. All the techniques mentioned in the paper are working around feature vectors or range images for most of the cases. All the datasets used in the presented researches had small 3D face datasets. In modern times, humungous size of datasets can be handled using deep learning techniques and advent of GPUs. Cho et al. [16] proposed a graph-structured module called Relational Graph Module (RGM), which focuses on the high-level relational information between the facial components. The heterogeneous face recognition (HFR) problem is handled in this work. HFR is a type of face recognition in which face is matched across two domains viz. near-infrared (NIR), visible light (VIS), or the sketch domain. RGM did the embedding of spatially correlated feature vectors into the graph node vectors and performs the relation modeling between different nodes of the graph. In addition to RGM, a Node Attention Unit (NAU) was used to perform node-wise recalibration. This model is able to handle HFR database.

Huang et al. [38] developed an adaptive curriculum learning loss (CurricularFace) for deep face recognition. CurricularFace addressed an idea of curriculum learning into a loss function for achieving a novel training technique. This technique addresses easy samples in the early training stage and hard samples in the later stage. Different importance is assigned to different samples based on the corresponding difficulty. The datasets used were CASIA-WebFace [84], refined MS1MS2 [18], LFW [47], CFP-FP [74], CPLFW [97], AgeDB [56], CALFW [98], IJB-B [80], IJB-C [53], and MegaFace [40]. Bi et al. [10] investigated the conditional GAN (cGAN) for understanding the face-to-sketch translation issues. Along with learning of mapping relationships between the face and the sketch, these networks generate a loss function for automatically training the mapping relationships. In the presented work, it is considered that multi-scale image representation can capture the structure, image texture, and other features accurately. Three layer pyramid model was constructed to obtain the multi-scale information. The multiscale cGAN model was used to train the mapping relationships. The datasets used were CUFS database [79], CUFSF dataset [88], and FERET database [65]. Fan et al. [27] presented a perceptual metric for facial sketches namely Structure Co-Occurrence Texture (Scoot). Scoot simultaneously considered the co-occurrence texture statistics and the block-level spatial structure.

Sharma and Kumar [75] presents a voxel based 3D face reconstruction technique using sequential deep learning. The datasets used in the presented work are Bosphorus, UMBDB, and Kinect Face DB. The process of voxelization is followed by variational autoencoders, bidirectional long short-term memory, and triplet loss training followed by support vector machine based prediction. The mirroring technique is used for reconstruction of the 3D voxelized face. Using the reconstructed face, a sequential deep learning framework is utilized for gender recognition, emotion type recognition, occlusion type recognition, and person identification.

Multiple deep learning metric algorithms [5, 6, 15, 59] have been designed loss function such that they can learn more distinguishing features. Evolutionary algorithms are mostly used for feature optimization because the search capability of these algorithms is better than others [83, 91]. In [21, 35, 36, 87, 92], the latest developments in machine learning, mathematical modeling, and optimization techniques are presented. The main shortcoming of the above-mentioned techniques is that it is difficult to recognize a face from 3D occluded face datasets. To resolve this problem, 3D occlusion invariant face recognition framework is proposed.

3 Proposed research framework

This section discusses the motivation followed by voxel-based 3D occlusion invariant face recognition framework.

3.1 Motivation

The proposed framework is motivated by the recent success of generative adversarial networks (GANs). The use of voxels makes it possible to include the finer details of 3D face. According to the best of author's knowledge, little work has been done in the field of voxels and deep learning for 3D face recognition. The proposed framework utilizes the concepts of voxelization, locality preserving projections, triplet loss, simulated annealing, and game theory. In the traditional approach, 3D mesh images are converted into depth images (2.5D) or Epipolar geometry-based multiple 2D images, which are used to train the conventional neural networks (CNNs) or autoencoders. In contrast to 2D and 2.5D images, the presented work is implemented using voxels in 3D form. In Sharma [75], mirroring technique based face reconstruction was done after voxelization along with BiLSTM based sequential deep learning. Figure 5 shows the comparison among traditional approach given by [22, 67], and the proposed approach of 3D face recognition framework. The proposed approach uses voxels in contrast to depth images or epipolar geometry images.

3.2 Proposed 3D face recognition framework

The proposed framework for 3D face recognition consists of two phases, namely, training and testing. Figure 6 presents the proposed 3D face recognition framework.

3.2.1 Training phase

There are two sub-phases in the training phase, namely, pre-processing and simulated annealing based deep learning. The detail descriptions of these phases are mentioned in preceding subsections.

3.2.2 Pre-processing

During the training phase, voxelization and locality preserving projections are two well-known preprocessing techniques for generating embeddings. Figure 7 shows the mesh images and their corresponding voxel images. The voxelization process converts a 3D mesh into voxel form in such a way that 3D coordinates are generated for each triangular mesh represented using cubes in different grid sizes. A single mesh is converted into three different voxel grid sizes viz. 4^3, 8^3, and 16^3. The number of voxels generated is sparse for different phases, even for the same size grid. Locality preserving projections are used to handle the problem of sparseness. 4^3 voxels are converted into 64×3

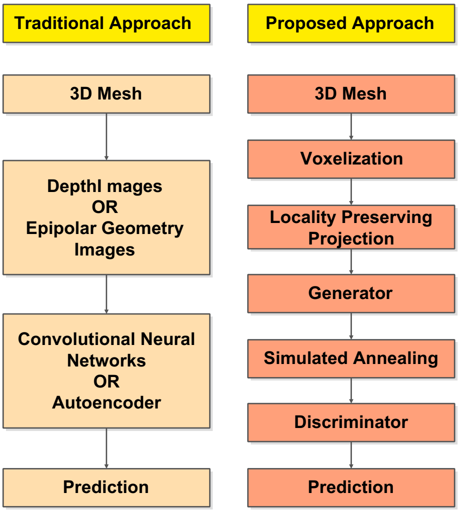

Fig. 5 Comparison between the traditional approach [22, 67] and proposed approach of 3D face recognition framework

embedding, 8^3 voxels are converted into 128×3 embedding, and 16^3 voxels are converted into 256×3 embedding. Ensembling is a famous technique in making the prediction model more robust towards new test images. Hence, three different kinds of grid sizes are used. It helps in boosting the quality of training data during the preprocessing step.

3.2.3 Adversarial voxel triplet generator and simulated annealing based prediction

The pre-processing sub-phase produces normalized voxel embedding for further processing. The generator produces triplets of Anchor (A), Positive (P), and Negative (N) for triplet loss training. Motivated from [95], normalized voxel embeddings for a voxelized mesh image x is represented as $V(x) \in \mathbb{R}^L$. Given <A, P, N> as a triplet, <A, P> is relevant (positive) pair and <A, N> is irrelevant (negative) pair. The objective function to train $V(x)$ such that minimizing the following loss:

$$L_{V,tri} = [d(V(a), V(p)) - d(V(a), V(n)) + m]_+ \tag{9}$$

where $d(x,y) = \left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\|^2$ is squared Euclidean distance between two L2-normalized vectors. m is the least margin required between d(a, p) and d(a, n) during training, and $[.]_+ \triangleq \max(., 0)$ denotes the positive component of the input. Let the adversarial voxel triplet generator (G) generates an adversarial sample $G(V(x)) \in \mathbb{R}^L$ by modifying the feature representation $V(x)$ of an image x. While generator training to minimize the triplet loss, G produces hard triplet examples by pushing away the same category vectors and pushing close the different category vectors.

The following objective is to be minimize the adversarial voxel triplet loss during training G,

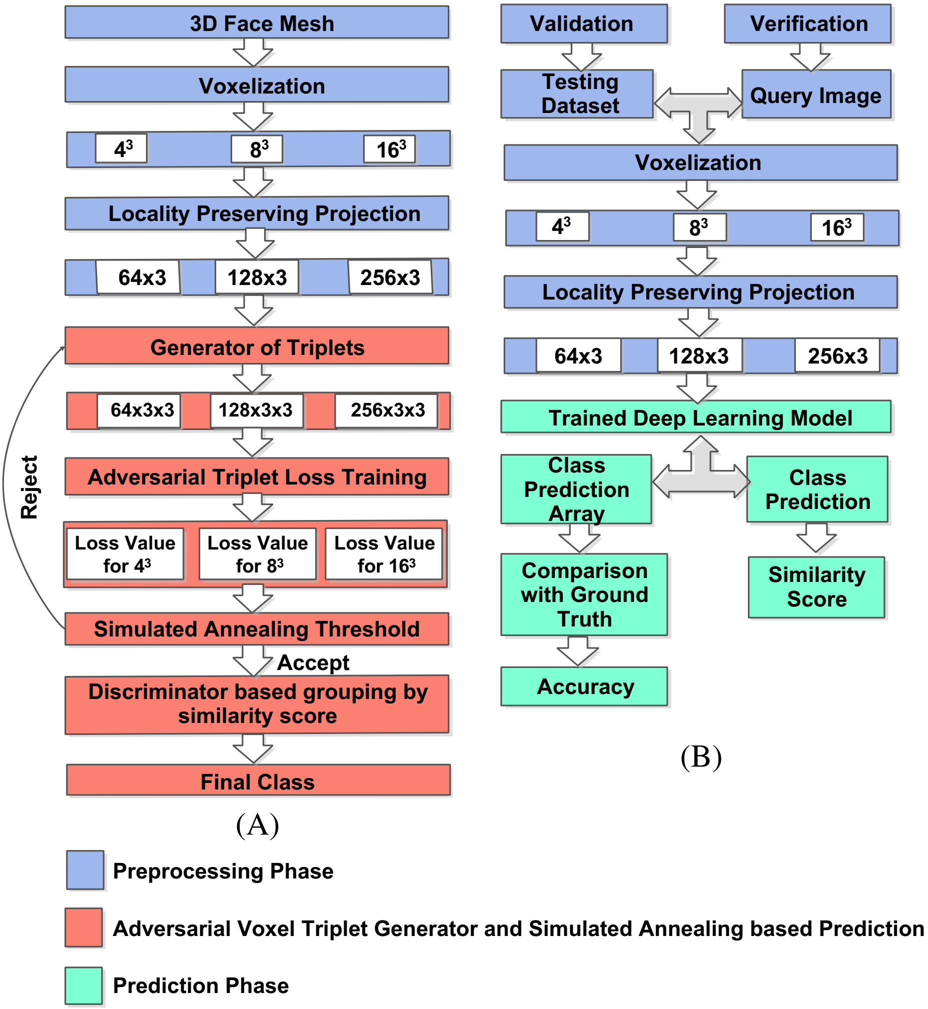

Fig. 6 Proposed framework for 3D face recognition (**a**) Training phase (**b**) Testing phase

$$L_{G,tri} = [d(G(V(a)), G(V(n))) - d(G(V(a)), G(V(p))) + m]_+ \tag{10}$$

Finally, with a fixed G objective function for training becomes

$$L_{V,tri} = [d(G(V(a)), G(V(p))) - d(G(V(a)), G(V(n))) + m]_+ \tag{11}$$

Here, $L_{G,\ tri}$ and $L_{V,\ tri}$ makes up an adversarial loss pair. Comparing Eq. (9) and Eq. (11), V is trained through the triplets generated by G pushing the <*A*, *P*> closer and <*A*, *N*> apart to meet margin m.

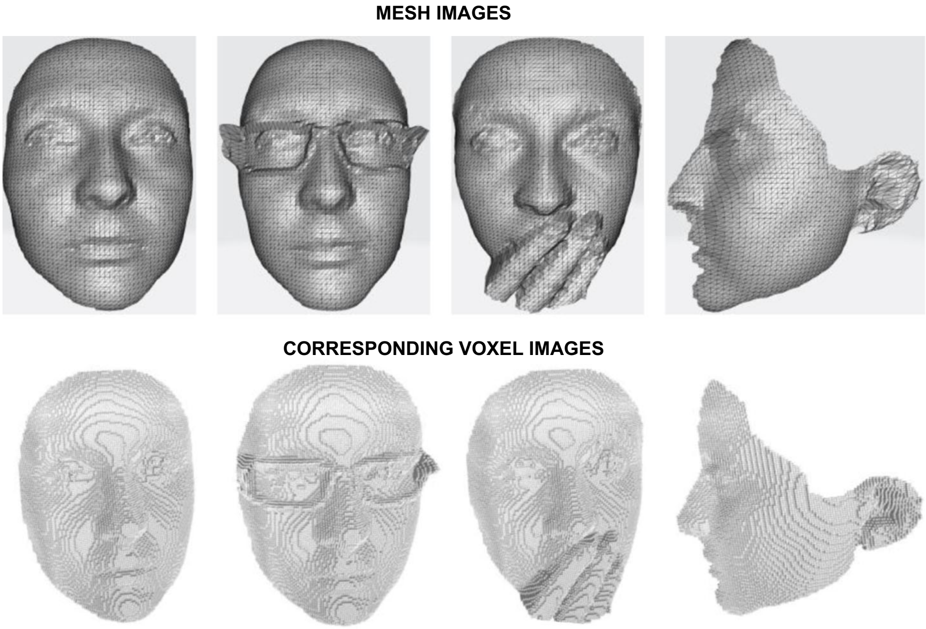

Fig. 7 Mesh images and its corresponding voxel images

The adversarial mechanism using a generator (G) is insufficient without the use of discriminator along with it. The role of discriminator (D) is to monitor and the constrain the triplet generator G from producing random triplet vectors for attaining a low value of $L_{G,\ tri}$. Using a discriminator D, a feature vector is categorized into (C + 1) categories, where real class examples are represented by the first C categories and the final one denotes the fake class. The triplet <A, P, N> has the labels<l_A, l_P, l_N>, the positive pair has $l_A = l_P$ and the negative pair has $l_A \neq l_N$. The following loss function is minimized for training D.

$$L_D = L_{D,real} + \beta L_{D,fake} \tag{12}$$

Here, D is forced to do the classification of feature vectors of the triplet correctly by the first term ($L_{D,\ real}$).

$$L_{D,real} = [L_{ll}(D(V(A)), l_A) + L_{ll}(D(V(P)), l_P) + L_{ll}(D(V(N)), l_N)]{*}0.33 \tag{13}$$

where L_{ll} signifies the log loss. However, the second term $L_{D,\ fake}$ enables D to differentiate between real features and the generated features.

$$L_{D,fake} = \left[L_{ll}\left(D(G(V(A))), l_{fake}\right) + L_{ll}\left(D(G(V(P))), l_{fake}\right) + L_{ll}\left(D(G(V(N))), l_{fake}\right)\right]{*}0.33 \tag{14}$$

Here, fake class is denoted by l_{fake}.

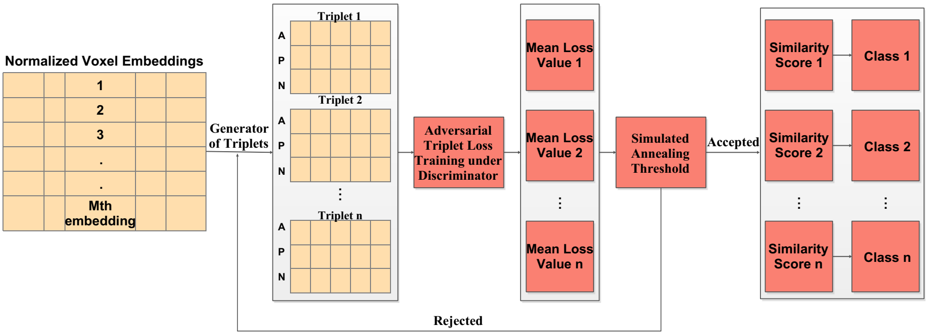

Fig. 8 Prediction of face using adversarial voxel triplet generator and simulated annealing

D plays a crucial role in helping *G* for the preservation of a class of input features. Hence, the subsequent loss enforces the class preservation assumption and represented as

$$L_{G,class} = [L_{ll}(D(G(V(A))), l_A) + L_{ll}(\mathrm{D}(G(V(P))), l_P) + L_{ll}(D(G(V(N))), l_N)]^{*}0.33 \quad (15)$$

The final loss value is minimized by training the voxel triplet generator *G* and defined as

$$L_G = L_{G,tri} + \gamma L_{G,class} \quad (16)$$

Based on the mean triplet loss for multiple grid sizes, simulated annealing threshold is applied for accepting the predicted similarity score. The concept of simulated annealing has been introduced here to make sure that the minimization problem of the mean loss value coming as output from adversarial triplet loss training under discriminator is handled in an effective way by keeping a check on the threshold value. If the mean loss value does not satisfy the simulated annealing threshold, then embedding is dropped and sent back to the generator for the new triplet generation. The similarity score and final class are generated through discriminator classifying the selected embeddings. Figure 8

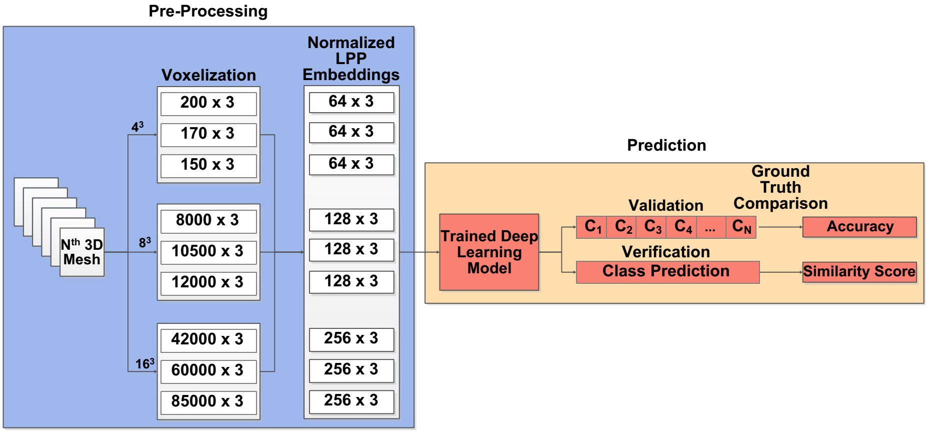

Fig. 9 Pre-processing and prediction processes in the testing phase

depicts prediction and score matching using adversarial triplet loss and simulated annealing. In this figure, M is the number of embeddings after the voxel normalization, and n is the number of triplets forming via generator.

3.2.4 Testing phase

There are two sub-phases in the testing phase, namely, pre-processing and the prediction for validation and verification. Figure 9 shows the pre-processing and prediction phase for different grid size voxels.

3.2.5 Pre-processing

The pre-processing during testing phase is considered either for validation or for verification at one time. For validation, the testing dataset is considered. Voxelization process is carried out on each image in the testing dataset. For verification, the voxelization process is carried out on a single query image. Locality preserving projection normalizes voxels removing their sparseness for deep learning model.

3.2.6 Prediction

In case of validation, an array of class predictions is given as output from trained deep learning model. For validation, the output array values are compared with ground truth values to calculate the accuracy of the model. In the case of verification, the predicted value is a single class. For verification, the final similarity score is calculated using the correlation value [49].

3.3 V3DOFR and computational complexity

The proposed algorithm of voxel-based 3D occlusion invariant face recognition (V3DOFR) consists of five steps. Firstly, raw 3D mesh image is taken as input, and the number of triangular units of mesh is counted. If there are no triangular units found, then an error message is generated (see Step 1). In Step 2, voxelization is performed for different grid sizes. The number of voxels and grid sizes are linearly proportional. During the process of voxelization, there is an inconsistency in voxel count due to different poses with in the same class. To overcome this inconsistency, locality preserving projection (LPP) is used in Step 3, that will help in removing the sparseness while maintaining the neighboring voxel properties. Thus, LPP is a more effective technique than principal component analysis (PCA) for dimensionality reduction in maintaining the voxel properties at facial feature level. Different grid sizes are converted into different number of LPP feature set. Once the LPP embeddings are generated, triplet generation is followed using generator in Step 4. The generator randomly selects Anchor (A), Positive (P), and Negative (N) embeddings. Deep learning-based triplet loss training is performed for computation of loss value. Further, normalization of loss values for corresponding grid sizes is performed in Step 4. In final step, average of normalized loss value is calculated for simulated annealing-based triplet selection. After triplet selection, discriminator assigns the class identification number. If triplet is not selected, then new triplet is generated through generator.

Algorithm. Proposed *V3DOFR* – Voxel based 3D Occlusion Invariant Face Recognition

```
Input: 3D Mesh Face Image (I)
Output: Classification of Image I
1.  Pre-processing of Mesh
        a.  Count traingular units of mesh ($C_t$)
        b.  If $C_t = 0$ do
        c.      Voxel = "Error: No triangular unit found"
        d.  End if
2.  Voxelization according to grid sizes of $4^3$, $8^3$, and $16^3$
        a.  Find the voxel for each triangular unit. Function: V(size)
        b.  For t = 1 to $C_t$ do
        c.      Using three coordinates A(x1,y1,z1), B(x2,y2,z2), and C(x3,y3,z3)
        d.      $X = (x1+x2+x3)/3$
        e.      $Y = (y1+y2+y3)/3$
        f.      $Z = (z1+z2+z3)/3$
        g.      Save $V_t = [X,Y,Z]$
        h.  End for
        i.  VoxelSize = [4, 8, 16];
        j.  $C_{v4}$ = [], $C_{v8}$ = [], $C_{v16}$ = [];      //Current Voxels for all sizes
        k.  $C_{v4} = V(4)$;
        l.  $C_{v8} = V(8)$;
        m.  $C_{v16} = V(16)$;
3.  Calculation of LPP embeddings
        a.  LPP4 = LPP($C_v$, 64);                    //LPP embedding of grid size $4^3$
        b.  LPP8 = LPP($C_v$, 128);                   //LPP embedding of grid size $8^3$
        c.  LPP16 = LPP($C_v$, 256);          //LPP embedding of grid size $16^3$
        d.  TotalLPP4 = [TotalLPP4;LPP4;];  //LPP embedding of grid size $4^3$ for all the dataset
        e.  TotalLPP8 = [TotalLPP8;LPP8;];  //LPP embedding of grid size $8^3$ for all the dataset
        f.  TotalLPP16 = [TotalLPP16;LPP16;];         //LPP embedding of grid size $16^3$ for all the dataset
        g.  TotalLPP = [TotalLPP4, TotalLPP8, TotalLPP16];
4.  Triplet Generation with Generator and Discriminator
        a.  Loss = [0,0,0]
        b.  NormLoss = [0,0,0]
        c.  For t = 1 to 3 do
        d.      Select embedding for A and P randomly from TotalLPP[t] of same class and N randomly from
                other class
        e.      Adversarial voxel triplet loss training
        f.      Loss[t] = triplet loss after training        //Logarithmic Loss value
        g.  End for
5.  Simulated Annealing based Prediction
        a.  FinalLoss = (Loss[1] + Loss[2] + Loss[3])/3;
        b.  Try
        c.      If($e^{FinalLoss}$ < rand(0,0.8)))      //Concept based on simulated annealing
                        //Accept
                        Throw($I_{ID}$)           //ID of the image
        d.      Else
                        //Reject
                        Goto: Step 4             //Go back to Generator for Triplet Generation
        e.      End If
        f.  End try
        g.  Catch($I_{ID}$)
                ID = Max (Corr($I_{ID}$)) among all classes
                FinalClass = ID
        h.  End catch
```

3.3.1 Computational complexity

The time complexity of the proposed algorithm is as follows. The pre-processing of mesh requires *O(n)* time. In *Step 2* voxelization (i.e. *2(a)-2(h))*, all steps require *O(n)* time and sub-steps (i.e. *2(i)-2(m))* requires *O(1)* time. In *Step 3* calculation of LPP embeddings (i.e. *3(a)-3(c))* requires $O(n^3)$ time [32], and other sub-steps (i.e. *3(d)-3(g))* requires *O(1)* time. In *Step 4* Triplet generation with generator and discriminator takes *O(1)* time for steps (i.e. *4(a)-4(b))* and $O(n^3)$ time [20] for rest of the sub-steps (i.e. *4(c)-4(g))*. The simulated annealing based prediction takes *O(1)* time. Hence, the total complexity of proposed technique is $O(n^3)$.

4 Experimental results and discussion

In this section, the performance of the proposed technique is compared with the existing techniques along with their visual verification. This section presents datasets used, parameter setting, and computational time analysis.

4.1 Datasets used

The datasets used for evaluation of the proposed techniques are Bosphorus face database [71], UMBDB face database [17], and KinectFaceDB face database [54]. Bosphorus dataset consists of 105 subjects in different poses and occlusions. There are 381 occluded images in Bosphorus dataset. All images are annotated with subject ID and pose, occlusion, or emotion description. The total number of images in Bosphorus dataset is 4666. For UMBDB dataset, there are 1473 different images. 590 images are occluded out of 1473 with a different type of occlusions. The number of subjects in the dataset is 143. The modalities of this dataset are 2D and 3D. In KinectFaceDB dataset, there are a total of 52 subject's data. Three types of modalities are covered in this dataset viz. 2D, 2.5D, and 3D. The total number of images is 936, and 312 images are occluded out of it. Table 1 presents the detail description of these datasets. Table 2 presents the occlusion description for these datasets.

4.2 Parameter setting

The parameters of the proposed approach are mentioned in Table 3. In the voxelization process, the grid sizes are kept as 4x4x4, 8x8x8, and 16x16x16, respectively. The corresponding number of neighbours for locality preserving projection are 16, 64, and 128 using the current voxels of the corresponding grid size. K-nearest neighbour along with adjacency weight matrix is assigned for an effective LPP embeddings. The number of epochs are 2700, 1200, and 800 corresponding to various grid sizes in triplet loss training. Adaptive moment (Adam) optimizer [43] is employed in triplet loss training. The alpha value is kept to be 0.2 and mean absolute error is used as a loss parameter. The loss function used in the discriminator is the logarithmic loss function, which directly gives the values in range 0 to 1. The batch size is kept to be 30, the dropout rate is kept at 40%, the learning rate is 0.005, and the activation function used is the rectified linear unit (ReLU).

Table 1 Description of datasets used

Dataset	Number of subjects	Number of images	Modality	Annotation
Bosphorus	105	4666	2D, 3D	Yes
UMBDB	143	1473	2D, 3D	Yes
KinectFaceDB	52	936	2D, 2.5D, 3D	Yes

Table 2 Occlusion description for datasets

Dataset	Bosphorus	KinectFaceDB	UMBDB
Occlusion Type Count	4	3	5
Occlusion Types	Eye, Mouth, Glass, Hair	Eye, Mouth, Paper	Cloth, Glass, Hair, Mouth, Paper
Occluded Images Count	381	312	590

ElSayed et al. [24] used Siamese neural network with (2, 500, 1) model, where 2 are the number of inputs, 500 are the number of nodes in the hidden layer and giving single output. Tan et al. [77] used ResNet-18 model with a 256 × 256 grid size of the depth map image. Adam optimizer has been used along with an initial learning rate of 0.01 and weight decay of 5×10^{-5}. Liu et al. [51] built a face reconstruction model based on the pose and expression normalization using 128 SIFT descriptors and tanh activation function for yaw poses of $0^{\circ}, \pm 10^{\circ}, \pm 20^{\circ}, \ldots, \pm 90^{\circ}$.

Table 3 Parameter setting

Parameter	Value
Proposed V3DOFR	
Voxelization	
Grid sizes	4^3, 8^3, 16^3
Locality preserving projection [33]	
Number of neighbours	16, 64, 128
Number of voxels	Current Voxels (V_C)
Neighbour algorithm	K-nearest neighbour
Weight	Adjacency
Adversarial Voxel Triplet loss training [95]	
Number of epochs	2700, 1200, 800
Optimizer	Adam
Alpha	0.2
Overall Loss	Mean Absolute Error
Discriminator Loss	Log Loss
Batch Size	30
Dropout	0.40
Learning Rate	0.005
Activation	ReLU
ElSayed et al. [24]	
Number of inputs	2
Number of hidden layers	1
Nodes in hidden layer	500
Number of outputs	1
Tan et al. [77]	
Input size	256 × 256
CNN model	ResNet-18
Optimizer	Adam
Weight decay	5×10^{-5}
Initial learning rate	0.01
Liu et al. [51]	
Number of SIFT descriptors	128
Yaw poses	$0^{\circ}, \pm 10^{\circ}, \pm 20^{\circ}, \ldots, \pm 90^{\circ}$
Activation	tanh

4.3 Performance evaluation metrics

The seven well-known performance evaluation measures such as accuracy, sensitivity, specificity, precision, FPR, FNR, and F1 score are used for comparing the quality of the proposed technique along with other techniques. These measures are computed from the confusion matrix and are shown in Fig. 10.

With reference to confusion matrix in Fig. 10, it is important to understand the concepts of true positive (TP), true negative (TN), false-positive (FP), and false-negative (FN). When actual and predicted values are 'YES', it is known as TP. When both values are 'NO', it is known as TN. In FN, the actual value is 'YES', but predicted is 'NO'. For FP, the actual value is 'NO', but predicted is 'YES'.

Accuracy is the measure of total correctness while predicting the classes and defined as [100].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

Sensitivity is the measure of correct classification of all the true positives and defined as [100].

$$Sensitivity = \frac{TP}{TP + FN} \tag{18}$$

Specificity is the measure of correct classification of all the true negatives and defined as [100].

$$Specificity = \frac{TN}{TN + FP} \tag{19}$$

Precision is defined as the ratio of actual positive values compared to total positive values including the predicted ones. It is mathematically represented as

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

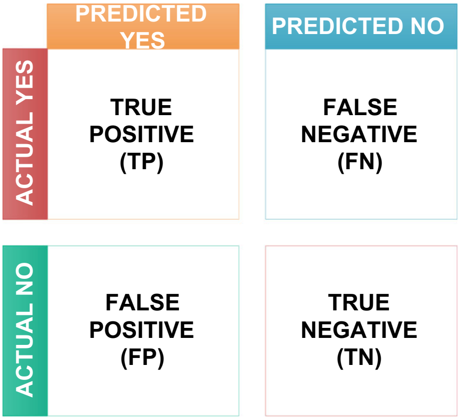

Fig. 10 Confusion matrix

	PREDICTED YES	PREDICTED NO
ACTUAL YES	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
ACTUAL NO	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

False Positive Ratio (FPR) is the ratio of wrongly predicted negative values to total number of negative values in actual and predicted. FPR is defined as

$$FPR = \frac{FP}{FP + TN} \tag{21}$$

False Negative Ratio (FNR) is the ratio of wrongly predicted positive values to total number of positive values in actual and predicted. The mathematical representation of FNR is as follows

$$FNR = \frac{FN}{FN + TP} \tag{22}$$

F1 Score is represented as the harmonic mean of precision and sensitivity value. It is defined as

$$F1\ Score = \frac{2*TP}{2*TP + FP + FN} \tag{23}$$

4.4 Non-adversarial versus adversarial voxel triplet generator face recognition technique

This section compares the techniques of 3D face recognition by using adversarial voxel triplet generator and without using the adversarial technique. Table 4 shows the performance comparison between non-adversarial and adversarial based voxel triplet generator.

The accuracy obtained over three datasets is 8–10% better in case of adversarial voxel-triplet based face recognition than the non-adversarial voxel-triplet generator based face recognition. Hence, the use of the adversarial technique in a combination of simulated annealing has proven to be beneficial for the computation of face recognition accuracy.

4.5 Performance evaluation

In this sub-section, the performance of the proposed techniques and four well-known techniques namely ElSayed [24], Tan [77], Liu [51], and Sharma [75] has been evaluated in four different experimentations. These are voxel, occlusion invariant face, landmarks, and 3D mesh. In each experimentation, the evaluation has been done through seven performance measures viz. Accuracy, Sensitivity, Specificity, Precision, False Positive Rate (FPR), False

Table 4 Comparison between proposed adversarial and non-adversarial based voxel triplet generator face recognition

Dataset	Accuracy		Sensitivity		Specificity	
	Adversarial voxel triplet generator	Non-adversarial voxel triplet generator	Adversarial voxel triplet generator	Non-adversarial voxel triplet generator	Adversarial voxel triplet generator	Non-adversarial voxel triplet generator
Bosphorus	90.8	82.7	95.8	94.7	70.0	32.7
UMBDB	81.9	75.4	88.5	90.2	43.8	44.5
Kinect face DB	85.6	77.2	92.7	92.4	45.4	34.8

Negative Ratio (FNR), and F1 Score. The validation of the proposed technique and compared algorithms have been tested over the dataset mentioned in Section 4.1.

Table 5 shows the performance comparison of various face recognition techniques using voxels. The training dataset has been generated using randomly selected 80% images from the given set. 20% images are used in the testing dataset. In Bosphorus dataset, the proposed technique provides better results than the existing techniques in terms of performance measures except specificity and FPR. While, Sharma [75] provides better value of specificity and FPR. The accuracy obtained from the proposed technique is 90.8%. Similarly, in UMBDB dataset, the proposed technique outperforms the other techniques in terms of performance measures except for sensitivity and specificity. The accuracy achieved by the proposed technique is 81.9%. The main reason behind to drop the accuracy in UMBDB dataset is that the presence of more dynamic occlusion present in UMBDB dataset as compared to Bosphorus dataset. In case of Kinect Face DB dataset, the best accuracy achieved through Sharma's method [75]. However, the proposed technique provides accuracy with a difference of 0.1%. Sharma [75] technique outperforms the others in terms of FNR and specificity. Whereas, precision, FPR, and F1 score obtained from the proposed technique is better than the existing techniques. The proposed technique and ElSayed [24] achieved the sensitivity at par with 92.7% and 92.9%, respectively.

Table 6 presents the results obtained from various face recognition techniques under occlusion environment. The proposed model and the other four techniques have been trained with the non-occluded images in the dataset. However, it has been tested on the occluded images. In Bosphorus dataset, the best accuracy obtained from the proposed technique is 81.5%. The accuracy achieved by the proposed approach is better than the second best technique by 2.1%. In terms of sensitivity, the proposed technique is the second best technique. For specificity, FPR, FNR, and F1 Score, the proposed method outperforms the other face recognition methods. In case of precision, the proposed technique and ElSayed [24] provide 84.1% and 85.3% value, respectively. In case of UMBDB and Kinect Face DB dataset, the proposed technique attained best value for all the performance measures except

Table 5 Performance measures obtained from various face recognition techniques using voxels

	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	F1 Score
Bosphorus Dataset							
ElSayed [24]	89.2	94.5	64.9	90.2	34.3	9.8	90.4
Tan [77]	87.4	95.3	50.2	89.4	54.6	13.1	88.1
Liu [51]	84.7	93.3	31.0	91.0	65.1	7.2	91.9
Sharma [75]	90.0	93.4	**73.7**	94.5	**26.3**	6.6	93.9
Proposed	**90.8**	**95.8**	70.0	**94.5**	28.4	**6.2**	**94.0**
UMBDB Dataset							
ElSayed [24]	79.2	**91.2**	36.3	87.9	59.2	14.3	87.2
Tan [77]	77.5	89.8	21.4	87.3	58.9	16.0	85.6
Liu [51]	72.8	80.5	**46.0**	80.4	73.7	12.5	83.8
Sharma [75]	78.2	89.7	31.6	84.2	68.4	10.3	86.8
Proposed	**81.9**	88.5	43.8	**88.7**	**53.2**	**8.8**	**89.5**
Kinect Face DB							
ElSayed [24]	82.8	**92.9**	27.1	89.5	48.8	10.4	89.6
Tan [77]	79.8	91.9	34.2	86.2	64.9	11.2	87.5
Liu [51]	74.3	84.4	31.3	85.7	57.5	17.5	84.1
Sharma [75]	**85.7**	92.2	**50.0**	91.0	50.0	**7.8**	91.6
Proposed	85.6	92.7	45.4	**92.6**	**41.6**	10.3	**92.2**

Best achieved values are highlighted

sensitivity. The best accuracies obtained from the proposed approach over UMBDB and Kinect Face DB are 67% and 77.9%, respectively. The sensitivity and specificity values achieved by using the proposed technique in case of UMBDB are 79.2% and 38.0%, and in case of Kinect Face DB are 88.1% and 40.4%, respectively.

Table 7 shows the results obtained from 3D face recognition techniques using landmarks. The training and testing datasets are generated in ratio of 80–20 randomly using 26 landmarks in each case. In case of Bosphorus dataset, the proposed technique has the best results for most of the performance measures. The highest accuracy achieved from the proposed face recognition technique is 84.9% with 93.8% sensitivity and 49.7% specificity. The proposed technique achieved precision as 90.8%, FPR as 36.6%, FNR as 8.3%, and F1 score as 91.3%. Sharma's method [75] provides the second best accuracy, sensitivity, FPR, FNR, and F1 score. In case of UMBDB dataset, the recognition accuracy obtained from the proposed approach is 77.4%. The proposed method outperforms the other methods in terms of performance measures except specificity. Sharma's method [75] provides better results than the proposed method in terms of specificity. In case of Kinect Face DB dataset, the proposed technique outperforms all the other methods in terms of evaluation metrics except specificity. Sharma's method [75] has more specificity than the proposed technique. The accuracy achieved from the proposed face recognition for Kinect Face DB is 81.6%.

Table 8 shows the results obtained from different face recognition techniques using mesh. The training and testing dataset is partitioned into 80–20 ratio randomly. While using Bosphorus dataset, the accuracy achieved from the proposed face recognition technique is 88.7%. While, Sharma [75] method provides 87.4%. The sensitivity achieved by the proposed technique is 92.8%. Tan [77] attained the best sensitivity value 94.1%. In all the other evaluation metrics, the proposed technique achieved the best results. In case of UMBDB dataset, the best accuracy for the 3D face recognition is achieved by the proposed technique with a value of 79.2%. The best sensitivity is achieved by ElSayed [24] with 5.2% difference from the proposed technique. The precision, FPR, FNR, and the F1 score of the proposed

Table 6 Performance measures obtained from the various face recognition techniques under occlusion condition

	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	F1 Score
Bosphorus Dataset							
ElSayed [24]	77.5	**93.7**	17.6	**85.3**	66.5	12.4	86.4
Tan [77]	75.6	91.2	42.4	83.9	54.0	15.5	84.2
Liu [51]	73.5	88.2	26.0	82.3	74.1	9.6	86.1
Sharma [75]	79.4	90.6	43.6	83.8	56.4	9.4	87.0
Proposed	**81.5**	93.2	**50.2**	84.1	**51.8**	**7.9**	**87.9**
UMBDB Dataset							
ElSayed [24]	64.2	83.9	25.5	72.5	81.3	22.8	78.2
Tan [77]	63.0	**86.7**	19.4	78.3	78.5	26.8	75.7
Liu [51]	60.9	79.1	16.9	79.1	83.5	28.3	75.2
Sharma [75]	65.6	74.6	37.7	78.8	62.6	25.4	76.7
Proposed	**67.0**	79.2	**38.0**	**79.9**	**62.3**	**15.2**	**78.6**
Kinect Face DB							
ElSayed [24]	74.6	**91.0**	24.9	80.0	62.6	14.8	82.5
Tan [77]	72.4	90.5	13.6	84.0	63.2	18.4	82.8
Liu [51]	71.3	89.7	19.2	78.0	76.3	12.7	82.4
Sharma [75]	76.7	85.9	39.1	83.9	58.9	14.1	85.1
Proposed	**77.9**	88.1	**40.4**	**85.0**	**56.2**	**9.6**	**85.4**

Best achieved values are highlighted

Table 7 Performance comparison between different 3D face recognition techniques using landmarks

	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	F1 Score
Bosphorus Dataset							
ElSayed [24]	83.4	90.6	28.8	90.4	46.2	9.9	90.2
Tan [77]	81.7	91.5	18.9	87.3	69.4	8.5	89.4
Liu [51]	77.5	85.8	55.8	90.3	67.0	13.8	88.2
Sharma [75]	84.6	92.4	28.5	88.4	41.5	8.3	90.3
Proposed	**84.9**	**93.8**	**49.7**	**90.8**	**36.6**	**5.6**	**91.3**
UMBDB Dataset							
ElSayed [24]	73.7	84.6	30.1	80.2	62.3	14.2	82.9
Tan [77]	72.3	83.7	34.8	81.3	78.3	14.8	83.2
Liu [51]	69.8	81.9	25.9	75.5	66.6	13.2	80.8
Sharma [75]	75.0	89.2	**35.7**	79.5	65.3	12.8	84.0
Proposed	**77.4**	**90.1**	29.7	**85.2**	**60.8**	**10.2**	**85.0**
Kinect Face DB							
ElSayed [24]	79.4	85.2	15.1	80.5	76.3	10.3	84.9
Tan [77]	77.1	83.3	27.6	88.0	61.0	16.8	85.5
Liu [51]	73.9	81.2	36.4	82.5	54.6	17.5	82.5
Sharma [75]	80.1	90.0	**37.9**	86.1	62.1	11.2	88.0
Proposed	**81.6**	**91.7**	19.2	**88.7**	**45.9**	**10.0**	**88.8**

Best achieved values are highlighted

technique are 86.6%, 48.7%, 10.4%, and 87.9% respectively. Liu [51] has performed better in case of specificity with 45.8% for UMBDB dataset. In case of Kinect Face DB, Sharma's method [75] outperforms the other methods including the proposed method for all evaluation metrics.

4.6 Visual verification

Visual verification of random 3D mesh images based on occlusion invariant proposed framework is presented in Fig. 11. All 3D meshes have an occlusion in them viz. hand on

Table 8 Performance comparison between different face recognition techniques using 3D mesh

	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	F1 Score
Bosphorus Dataset							
ElSayed [24]	87.1	93.5	63.6	91.1	34.6	7.3	92.7
Tan [77]	85.9	**94.1**	49.2	91.6	35.4	6.0	92.9
Liu [51]	81.3	90.1	32.4	87.7	58.7	8.9	89.3
Sharma [75]	87.4	93.8	38.3	92.2	61.8	6.24	92.9
Proposed	**88.7**	92.8	**68.0**	**92.8**	**30.3**	**5.8**	**93.4**
UMBDB Dataset							
ElSayed [24]	78.4	**90.4**	34.4	83.0	49.9	13.4	86.4
Tan [77]	75.2	88.5	20.5	84.0	60.7	13.0	86.7
Liu [51]	70.6	79.3	**45.8**	78.3	65.0	12.5	82.7
Sharma [75]	77.4	89.5	34.9	82.8	65.1	10.5	86.0
Proposed	**79.2**	85.2	42.7	**86.6**	**48.7**	**10.4**	**87.9**
Kinect Face DB							
ElSayed [24]	80.7	89.0	25.8	86.6	48.8	8.7	89.7
Tan [77]	77.5	90.7	33.7	88.9	45.7	14.9	86.6
Liu [51]	72.3	82.1	30.2	82.4	56.7	9.6	86.2
Sharma [75]	**85.7**	**92.1**	**58.4**	**90.4**	**41.6**	**7.9**	**91.2**
Proposed	83.7	90.1	42.3	87.5	51.4	12.9	87.3

Best achieved values are highlighted

eyes, hair, glasses, hands-on mouth, cloth, cap, and finger. Ten 3D meshes from the occluded images are selected randomly for verification. Predicted subject IDs are given along with actual subject IDs. Nine out of ten meshes have the correct predicted value during verification. Hence, it validates that the proposed method is an occlusion invariant.

4.7 Computational time analysis

Table 9 depicts GPU based computational time obtained from the proposed approach and other techniques. Four well-known 3D face recognition techniques are compared with the proposed method using voxels, landmarks, and meshes for pre-processing, recognition, verification, and the corresponding learning model. The computation time presented is calculated as the average time in all the phases. This work has been run on GeForce GTX 1080Ti GPU model with 3584 CUDA (Compute Unified Device Architecture) cores and a memory speed of 11 Gbps.

The overall time computation of proposed technique using the landmarks is the fastest and using meshes the slowest when compared to the proposed technique using the voxels.

4.8 Convergence analysis

The convergence of accuracy obtained from the proposed technique for all datasets is shown in Fig. 12. Bosphorus dataset converges at 90% accuracy in 2700 epochs, UMBDB dataset converges at 81% accuracy in 1200 epochs, and KinectFaceDB dataset converges at 85% in 800 epochs. The convergence plot is based on accuracy obtained from combined approaches of triplet loss training, simulated annealing, and game theory.

Figure 13 shows the accuracies obtained from the face recognition model using voxelized technique on three datasets viz. Bosphorus, UMBDB, and KinectFaceDB. It can be seen from

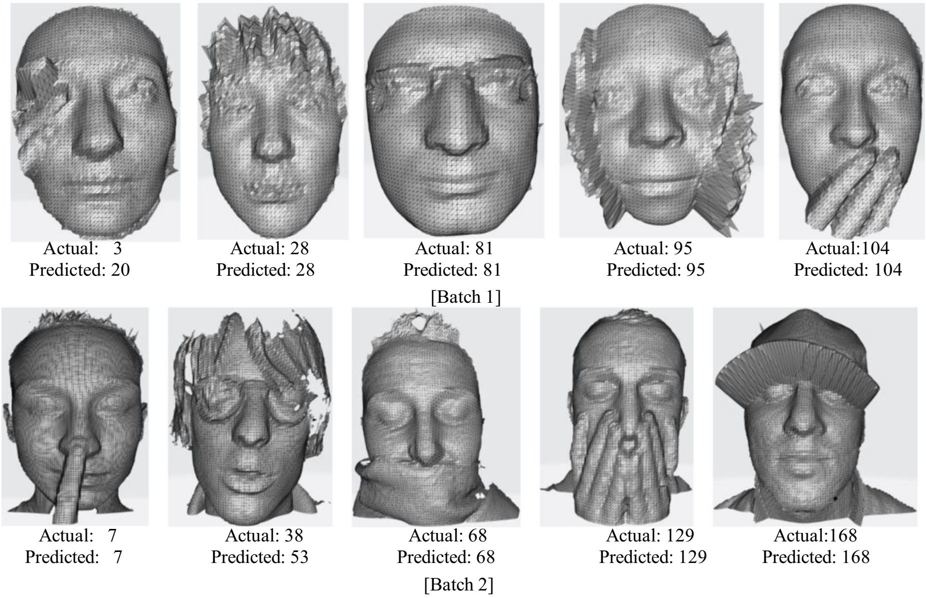

Fig. 11 Visual verification of 3D meshes with their actual and predicted subject IDs in Batch 1 and Batch 2

Table 9 Computation time (in ms) on GPU for proposed technique versus other techniques in voxel-based face recognition

Technique	Pre-Processing	Recognition	Verification	Learning Model
ElSayed et al. [24]	90	81	107	Siamese neural network
Tan et al. [77]	87	88	104	Convolutional neural network
Liu et al. [51]	96	73	112	Pose and expression normalized
Sharma et al. [75]	78	66	98	VAE, BiLSTM
Proposed using voxels	80	65	99	Adversarial triplet loss
Proposed using landmarks	55	67	95	Adversarial triplet loss
Proposed using meshes	105	69	105	Adversarial triplet loss

Fig. 13 that the accuracy on Bosphorus dataset is 90.8%, UMBDB is 81.9%, and KinectFaceDB dataset is 85.6%. These values can be verified in Table 5 performance measures obtained from various face recognition techniques using voxels.

5 Future work

The attention based models are being used for improving the accuracy of facial expression recognition [39]. There are other attention based models viz. image based attention [26], edge based attention [96], weakly supervised attention [93], and uncertainty based attention [94]. In [26], a high quality dataset namely SOC (Salient Objects in Clutter) is used to update the previous saliency benchmark for salient object detection. The attention of the deep learning model is brought to the objects in the image and the target is to detect the salient object in clutter and bring it to the foreground. This technique can be extremely useful to detect the facial landmarks such as eyes behind eyeglasses. The facial features can be effectively reconstructed using this approach by bring the salient facial features in the foreground of the occluding object. In [96], EGNet based on edge guidance network is presented for salient object detection. It focuses on the complementarity of salient edge information and salient object information to generate fine boundaries. This technique can be used in face detection similar to shape-from-shadow technique. The shape from shadow as well as shape from fully convolutional neural networks (FCNs) suffers from coarse object boundaries. Due to rich edge information, the salient objects can be detected more precisely. Hence, the facial features can be detected more precisely with fine edges using EGNet.

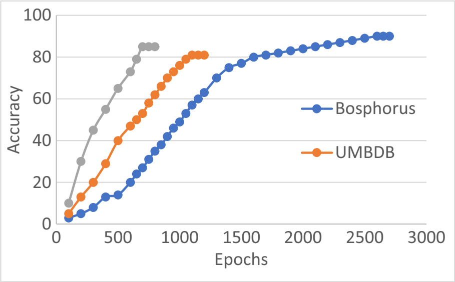

Fig. 12 Accuracy based convergence plot for proposed approach

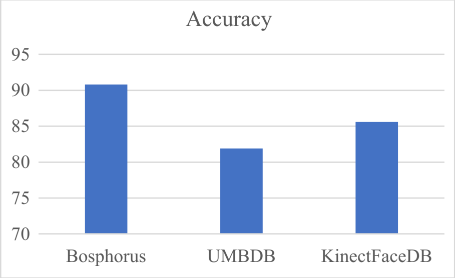

Fig. 13 Final accuracy comparison of three datasets

In [93], the labeling based salient object detection is proposed using weak-supervision technique. However, there is a challenge of poor boundary localization. To handle this problem, an auxiliary edge detection task is suggested for localization of object edges explicitly. This technique can be extended to use in 3D face detection for localization of facial features such as eyes, nose, mouth, etc. In [94], uncertainty inspired RGB-D saliency detection via conditional variational autoencoders (UC-Net) is presented. A probabilistic RGB-D saliency detection network is developed using conditional variational autoencoders for modeling of human annotation and build various saliency maps for each input image by latent space sampling. This technique can be used on RGB-D images for facial landmarks detection and predicting the facial expression.

The above-mentioned techniques may be utilized in the proposed approach for better performance in near future. EGNet can be integrated with the proposed approach for better facial feature extraction. RGB-D saliency detection method can be used in the proposed approach for landmark identification and detection. The attention based models may be utilized in the proposed approach for better recognition.

The simulated annealing based deep learning techniques can be implemented in all types of CNN models where backward propagation is done to calculate the loss between different layers. This simulated annealing can also be used in other deep learning models viz. autoencoders, variational autoencoders, GANs etc. because in all of them a simulated annealing based threshold value can be kept for loss acceptance.

6 Conclusions

In this paper, voxel-based 3D occlusion invariant face recognition framework is proposed. The proposed framework utilizes the concept of generator and discriminator based deep learning. Bosphorus, UMBDB, and Kinect Face DB have been used for implementing face recognition techniques. The best average accuracy obtained from face recognition using voxels by the proposed technique is 86.1%. Similarly, for occlusion invariant face recognition, the best average given by the proposed technique is 75.5%. In case of face recognition using 3D landmarks, the best average accuracy for the proposed technique is 81.3%. In case of face recognition using 3D meshes the best average accuracy given by the proposed technique is 83.9%. Adding the adversarial training strategy for triplet generation ensures low biasness. This technique, coupled with simulated annealing allows the proposed method to be robust in different areas using voxels.

References

1. Abrevaya VF, Boukhayma A, Wuhrer S and Boyer E(2019) A decoupled 3D facial shape model by adversarial training. In proceedings of the IEEE international conference on computer vision (pp. 9419-9428).
2. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S. and Asari, V.K. (2018) The history began from AlexNet: a comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164
3. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, Van Essen BC, Awwal AA, Asari VK (2019) A state-of-the-art survey on deep learning theory and architectures. Electronics 8(3): 292
4. Antipov G, Baccouche M and Dugelay JL (2017) Face aging with conditional generative adversarial networks. In 2017 IEEE International Conference on Image Processing (ICIP). pp. 2089–2093. IEEE
5. Bai S, Zhou Z, Wang J, Bai X, Jan Latecki L, Tian Q (2017) Ensemble diffusion for retrieval. In Proceedings of the IEEE International Conference on Computer Vision. pp. 774–783
6. Bai S, Bai X, Tian Q, Latecki LJ (2017) Regularized diffusion process for visual retrieval. In Thirty-First AAAI Conference on Artificial Intelligence
7. Bandyopadhyay S, Maulik U, Pakhira MK (2001) Clustering using simulated annealing with probabilistic redistribution. Int J Pattern Recognit Artif Intell 15(02):269–285
8. Bandyopadhyay S, Saha S, Maulik U, Deb K (2008) A simulated annealing-based multiobjective optimization algorithm: AMOSA. IEEE Trans Evol Comput 12(3):269–283
9. Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. In Advances in neural information processing systems. pp. 585–591
10. Bi H, Li N, Guan H, Lu D and Yang L (2019, September) A multi-scale conditional generative adversarial network for face sketch synthesis. In 2019 IEEE international conference on image processing (ICIP) (pp. 3876-3880). IEEE.
11. Bowyer KW, Chang K, Flynn P (2004) A survey of 3D and multi-modal 3D+ 2D face recognition.
12. CASIA-3D FaceV1, 3d face database
13. Caves R, Quegan S, White R (1998) Quantitative comparison of the performance of SAR segmentation algorithms. IEEE Trans Image Process 7(11):1534–1546
14. Chen Y, Garcia EK, Gupta MR, Rahimi A, Cazzanti L (2009) Similarity-based classification: concepts and algorithms. J Mach Learn Res 10:747–776
15. Chen W, Chen X, Zhang J and Huang K (2017) Beyond triplet loss: a deep quadruplet network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 403–412
16. Cho M, Kim T, Kim IJ and Lee S (2020) Relational deep feature learning for heterogeneous face recognition. arXiv preprint arXiv:2003.00697.
17. Colombo A, Cusano C, Schettini R (2011) UMB-DB: A database of partially occluded 3D faces. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). pp. 2113–2119. IEEE
18. Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4690-4699).
19. Ding C, Tao D (2016) A comprehensive survey on pose-invariant face recognition. ACM Transactions on intelligent systems and technology (TIST) 7(3):1–42
20. Do TT, Tran T, Reid I, Kumar V, Hoang T, Carneiro G (2019) A theoretically sound upper bound on the triplet loss for improving the efficiency of deep distance metric learning. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 10404-10413).
21. Dong Y, Zhang Z, Hong WC (2018) A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. Energies 11(4):1009
22. Dou P, Shah SK and Kakadiaris IA (2017) End-to-end 3D face reconstruction with deep neural networks. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5908-5917).
23. El Sayed AR, El Chakik A, Alabboud H, Yassine A (2018) Efficient 3D point clouds classification for face detection using linear programming and data mining. The Imaging Science Journal 66(1):23–37
24. El Sayed A, Kongar E, Mahmood A, Sobh T and Boult T (2018) Neural generative models for 3D faces with application in 3D texture free face recognition. arXiv preprint arXiv:1811.04358
25. Faltemier TC, Bowyer KW and Flynn PJ (2007) Using a multi-instance enrollment representation to improve 3D face recognition. In 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems. pp. 1–6. IEEE

26. Fan DP, Cheng MM, Liu JJ, Gao SH, Hou Q and Borji A (2018) Salient objects in clutter: bringing salient object detection to the foreground. In proceedings of the European conference on computer vision (ECCV) (pp. 186-202).
27. Fan DP, Zhang S, Wu YH, Liu Y, Cheng MM, Ren B, Rosin PL and Ji R (2019) Scoot: a perceptual metric for facial sketches. In proceedings of the IEEE international conference on computer vision (pp. 5612-5622).
28. Gecer B, Ploumpis S, Kotsia I and Zafeiriou S (2019) GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction. arXiv preprint arXiv:1902.05978
29. Goodfellow I (2016) NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv: 1701.00160
30. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In Advances in neural information processing systems:2672–2680
31. Hassaballah M, Aly S (2015) Face recognition: challenges, achievements and future directions. IET Comput Vis 9(4):614–626
32. He X (2005) Locality preserving projections. The University of Chicago, A dissertation submitted to the faculty of the division of the physical sciences in candidacy for the degree of doctor of philosophy Department of Computer Science
33. He X, Niyogi, P (2004) Locality preserving projections. In Advances in neural information processing systems pp. 153–160
34. He Z, Zuo W, Kan M, Shan S, Chen X (2019) Attgan: facial attribute editing by only changing what you want. IEEE Trans Image Process 28:5464–5478
35. Hong WC, Dong Y, Lai CY, Chen LY, Wei SY (2011) SVR with hybrid chaotic immune algorithm for seasonal load demand forecasting. Energies 4(6):960–977
36. Hong WC, Li MW, Geng J, Zhang Y (2019) Novel chaotic bat algorithm for forecasting complex motion of floating platforms. Appl Math Model 72:425–443
37. Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process 5(2):1
38. Huang Y, Wang Y, Tai Y, Liu X, Shen P, Li S, Li J and Huang F (2020) CurricularFace: adaptive curriculum learning loss for deep face recognition. arXiv preprint arXiv:2004.00288.
39. Jiao Y, Niu Y, Zhang Y, Li F, Zou C, Shi G (2019, December) Facial attention based convolutional neural network for 2D+ 3D facial expression recognition. In 2019 IEEE visual communications and image processing (VCIP) (pp. 1-4). IEEE.
40. Kemelmacher-Shlizerman I, Seitz SM, Miller D and Brossard E (2016) The megaface benchmark: 1 million faces for recognition at scale. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4873-4882).
41. Kim D, Hernandez M, Choi J and Medioni G (2017) Deep 3D face identification. In 2017 IEEE International Joint Conference on Biometrics (IJCB) pp. 133–142. IEEE
42. Kim D, Hernandez M, Choi J, Medioni G (2017) Deep 3D face identification. In 2017 IEEE International Joint Conference on Biometrics (IJCB). pp. 133–142. IEEE
43. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980
44. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. science, 220(4598), pp.671–680
45. Korshunov P and Marcel S (2018) DeepFakes: a new threat to face recognition? Assessment and Detection. arXiv preprint arXiv:1812.08685
46. Larsen, A.B.L., Sønderby, S.K., Larochelle, H. & Winther, O. (2016) Autoencoding beyond pixels using a learned similarity metric. Proceedings of The 33rd International Conference on Machine Learning, in PMLR 48:1558–1566
47. Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H. and Hua, G., 2016. Labeled faces in the wild: a survey. In advances in face detection and facial image analysis (pp. 189-248). Springer, Cham.
48. Lei Y, Guo Y, Hayat M, Bennamoun M, Zhou X (2016) A two-phase weighted collaborative representation for 3D partial face recognition with single sample. Pattern Recogn 52:218–237
49. Li H, Huang D, Morvan JM, Chen L, Wang Y (2014) Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. Neurocomputing 133: 179–193
50. Li H, Huang D, Morvan JM, Wang Y, Chen L (2015) Towards 3D face recognition in the real: a registration-free approach using fine-grained matching of 3D keypoint descriptors. Int J Comput Vis 113(2):128–142
51. Liu F, Zhao Q, Zeng D (2018) Joint face alignment and 3D face reconstruction with application to face recognition. IEEE Trans Pattern Anal Mach Intell

52. Maulik U, Bandyopadhyay S, Trinder JC (2001) SAFE: an efficient feature extraction technique. Knowl Inf Syst 3(3):374–387
53. Maze B, Adams J, Duncan JA, Kalka N, Miller T, Otto C, Jain AK, Niggel WT, Anderson J, Cheney J and Grother P, (2018, February) Iarpa janus benchmark-c: face dataset and protocol. In 2018 international conference on biometrics (ICB) (pp. 158-165). IEEE.
54. Min R, Kose N, Dugelay JL (2014) Kinectfacedb: a kinect database for face recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems 44(11):1534–1548
55. Moreno A (2004) GavabDB: a 3D face database. In Proc. 2nd COST275 workshop on biometrics on the internet, 2004 (pp. 75-80).
56. Moschoglou S, Papaioannou A, Sagonas C, Deng J, Kotsia I, Zafeiriou S (2017) Agedb: the first manually collected, in-the-wild age database. In proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 51-59).
57. ND-2006 Face Data Set. http://www.nd.edu/˜cvrl/. 2007.
58. Ogáyar CJ, Rueda AJ, Segura RJ, Feito FR (2007) Fast and simple hardware accelerated voxelizations using simplicial coverings. Vis Comput 23(8):535–543
59. Oh Song H, Xiang Y, Jegelka S, Savarese S (2016) Deep metric learning via lifted structured feature embedding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4004–4012
60. Pantaleoni J (2011) VoxelPipe: a programmable pipeline for 3D voxelization. In Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics. pp. 99–106. ACM
61. Parkhi OM, Vedaldi A, Zisserman A (2015) Deep face recognition. In British Machine Vision Conference (BMVC) 1(3):6
62. Patil H, Kothari A, Bhurchandi K (2015) 3-D face recognition: features, databases, algorithms and challenges. Artif Intell Rev 44(3):393–441
63. Perarnau G, Van De Weijer J, Raducanu B and Álvarez JM (2016) Invertible conditional gans for image editing. arXiv preprint arXiv:1611.06355
64. Pham HX, Chen C, Dao LN, Pavlovic V, Cai J and Cham TJ (2015) Robust performance-driven 3d face tracking in long range depth scenes. arXiv preprint arXiv:1507.02779
65. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
66. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005, June) Overview of the face recognition grand challenge. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 947-954). IEEE.
67. Ranjan A, Bolkart T, Sanyal S, Black MJ (2018) Generating 3D faces using convolutional mesh autoencoders. In proceedings of the European conference on computer vision (ECCV) (pp. 704-720).
68. Rathgeb C, Dantcheva A, Busch C (2019) Impact and detection of facial beautification in face recognition: an overview. IEEE Access 7:152667–152678
69. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A and Chen X (2016) Improved techniques for training gans. In Advances in neural information processing systems. pp. 2234–2242
70. Sanderson C (2002) The vidtimit database (No. REP_WORK). IDIAP
71. Savran A, Alyüz N, Dibeklioğlu H, Çeliktutan O, Gökberk B, Sankur B and Akarun L (2008) Bosphorus database for 3D face analysis. In European Workshop on Biometrics and Identity Management. pp. 47–56. Springer, Berlin, Heidelberg
72. Scherhag U, Rathgeb C, Merkle J, Breithaupt R, Busch C (2019) Face recognition systems under morphing attacks: a survey. IEEE Access 7:23012–23026
73. Schroff F, Kalenichenko D and Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823
74. Sengupta S, Chen JC, Castillo C, Patel VM, Chellappa R and Jacobs DW (2016, March) Frontal to profile face verification in the wild. In 2016 IEEE winter conference on applications of computer vision (WACV) (pp. 1-9). IEEE.
75. Sharma S, Kumar V (2020) Voxel-based 3D face reconstruction and its application to face recognition using sequential deep learning. Multimedia tools and applications, pp.1-28.
76. Spreeuwers L (2011) Fast and accurate 3d face recognition. Int J Comput Vis 93(3):389–414
77. Tan Y, Lin H, Xiao Z, Ding S and Chao H (2018) Face recognition from sequential sparse 3D data via deep registration. arXiv preprint arXiv:1810.09658
78. Vijayan V, Bowyer KW, Flynn PJ, Huang D, Chen L, Hansen M, Ocegueda O, Shah SK, Kakadiaris IA (2011, October) Twins 3D face recognition challenge. In 2011 international joint conference on biometrics (IJCB) (pp. 1-7). IEEE.

79. Wang X, Tang X (2008) Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell 31(11):1955–1967
80. Whitelam C, Taborsky E, Blanton A, Maze B, Adams J, Miller T, Kalka N, Jain AK, Duncan JA, Allen K and Cheney J (2017) Iarpa janus benchmark-b face dataset. In proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 90-98).
81. Wu Z, Song S, Khosla A, Tang X, Xiao J (2014) 3D Shapenets for 2.5D object recognition and next-best-view prediction. arXiv preprint arXiv:1406.5670, 2(4)
82. Xu D, Hu P, Cao W, Li H (2008, June) SHREC'08 entry: 3D face recognition using moment invariants. In 2008 IEEE international conference on shape modeling and applications (pp. 261-262). IEEE.
83. Yang XS (2010) Nature-inspired metaheuristic algorithms. Luniver press
84. Yi D, Lei Z, Liao S, Li SZ (2014) Learning face representation from scratch. arXiv preprint arXiv:1411.7923.
85. Yin L, Wei X, Sun Y, Wang J, Rosato MJ (2006, April) A 3D facial expression database for facial behavior research. In 7th international conference on automatic face and gesture recognition (FGR06) (pp. 211-216). IEEE.
86. Yin L, Sun\ XCY, Worm T and Reale M (2008) A high-resolution 3d dynamic facial expression database. In IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands. 126
87. Zhang Z, Hong WC (2019) Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. Nonlinear Dynamics 98(2):1107–1136
88. Zhang W, Wang X, Tang X (2011, June) Coupled information-theoretic encoding for face photo-sketch recognition. In CVPR 2011 (pp. 513-520). IEEE.
89. Zhang X, Yin L, Cohn JF, Canavan S, Reale M, Horowitz A and Liu P (2013, April) A high-resolution spontaneous 3d dynamic facial expression database. In 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG) (pp. 1-6). IEEE.
90. Zhang X, Yin L, Cohn JF, Canavan S, Reale M, Horowitz A, Liu P, Girard JM (2014) Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image Vis Comput 32(10):692–706
91. Zhang Y, Zhang L, Neoh SC, Mistry K, Hossain MA (2015) Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles. Expert Syst Appl 42(22): 8678–8697
92. Zhang Z, Hong WC, Li J (2020) Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm. IEEE Access 8:14642–14658
93. Zhang J, Yu X, Li A, Song P, Liu B and Dai Y (2020) Weakly-supervised salient object detection via scribble annotations. arXiv preprint arXiv:2003.07685.
94. Zhang J, Fan DP, Dai Y, Anwar S, Saleh FS, Zhang T and Barnes N (2020) UC-net: uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. arXiv preprint arXiv:2004.05763.
95. Zhao Y, Jin Z, Qi GJ, Lu H and Hua XS (2018) An adversarial approach to hard triplet generation. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 501–517
96. Zhao JX, Liu JJ, Fan DP, Cao Y, Yang J and Cheng MM (2019) EGNet: edge guidance network for salient object detection. In proceedings of the IEEE international conference on computer vision (pp. 8779-8788).
97. Zheng T, Deng W, (2018) Cross-pose lfw: a database for studying cross-pose face recognition in unconstrained environments. Beijing University of Posts and Telecommunications, Tech. Rep, 5.
98. Zheng T, Deng W, Hu J (2017) Cross-age lfw: a database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197.
99. Zhou Y, Deng J, Kotsia I and Zafeiriou S (2019) Dense 3D face decoding over 2500FPS: Joint Texture & Shape Convolutional Mesh Decoders. arXiv preprint arXiv:1904.03525
100. Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: health care and life sciences, Baltimore, Maryland 19:67
101. Zulqarnain Gilani S, Mian A (2018) Learning from millions of 3d scans for large-scale 3d face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 1896-1905