3D face construction from single 2D images using DEO model

Sumathi Doraikannan (sumathi.research28@gmail.com)
Vellore Institute of Technology - Amaravati Campus: VIT-AP Campus https://orcid.org/0000-0003-2920-4640
Varanasi LVSKB Kasyap
Vellore Institute of Technology - Amaravati Campus: VIT-AP Campus
Mure Sai Jaideep Reddy
Vellore Institute of Technology - Amaravati Campus: VIT-AP Campus
Varanasi Srinivasa Bhagavan
KLEF: KL Deemed to be University
Thangamuthu Poongodi
Galgotias University
Thangamariappan GANESH KUMAR
Galgotias University
Santhosh Kumar SVN
VIT University: Vellore Institute of Technology

Research Article

Keywords: CNN, deep learning, face construction, face parsing, facial geometry

Posted Date: May 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2658152/v1

3D face construction from single 2D images using DEO model

Doraikannan Sumathi, Varanasi LVSKB Kasyap, Mure Sai Jaideep Reddy, Varanasi Srinivasa Bhagavan, Thangamuthu Poongodi, Thangamariappan Ganesh Kumar, Santhosh Kumar SVN

***Abstract*—In recent years, considerable attention has been paid to 3D face data in many face image processing applications. Detailed 3D Face making is developing technology with multiple real-time applications. This work aims to create an exact 3D Face model with facial emotions designed based on the principle of the Face Vertex Land marking and Wulcheir distance. Convolution Neural Network (DCNN) is deployed to extract relevant facial features and those features are used for further analysis. The 3D Face models are constructed efficiently. The proposed model is a concoction of CoarseNet and FineNet through which a 3D coarse face from a bilinear face model with face landmark alignment is created. It is followed by the local corrective field which tends to refine the 3D rough face with consistent photometric constraint. This work follows the various aspects of 3D face modeling techniques: Deep Learning, Epiploic Geometry, and the One-shot learning (DEO) method. The proposed DEO Model has been evaluated using the FER2013 dataset of face images with six basic emotions via performance metrics like accuracy, precision, sensitivity, specificity, and time. The proposed model outperforms other existing methods with promising and state-of-art results. The accuracy obtained through the proposed work shows higher accuracy (more than 90%), which has been demonstrated using real-world models**

***Index Terms*— CNN, deep learning, face construction, face parsing, facial geometry.**

I. INTRODUCTION

COMPLEX expressions on the face make us explicitly express, and communications could still be made better through various modes like gestures, mimics and moments, etc., Among the various research problems in computer vision and video processing, detailed 3D face-making has attracted many researchers. However, a significant amount of loss during the processing of the 2D face image projections depends on textures, material properties, lighting conditions, and view directions. The flexibility of the face makes the facial geometry recovery from a single 2D image a challenge. Many methods have been deployed to construct 3D face model from single images. Amongst all the other methods, low-dimensional constraint representation of images is built by example-based methods and then fits into the parametric model of 2D image input. 3D Morphable Models (3DMM) are based on principal component analysis that highly uses example-based models for generating 3D Face models, illustrated as the linear combination of faces. Shape-from-shading (SFS) is an another approach for constructing 3D face models. The proposed methods can produce high-quality 3D face models, but the results are unsatisfied when target faces are different from the example set. The proposed models heavily rely on the dataset. The geometric details of faces are target-specific and are not reproduced because of a lesser degree of randomness in the low dimensionality model. SFS-based techniques can reproduce the faces with the fine-scale details when they have prior knowledge of the geometric features, and the produced faces may be inaccurate when the input image is less scaled. The novelty of this work is found in recommending the coarse-fine method to construct a detailed and face of superior quality that can be obtained from a specific and distinct 2D image. The proposed method is divided into three parts:

- Coarse estimation of the 3D face is computed from an input image fitted with an example-based parametric face. The parametric model used here is derived from two 3D face datasets, FACE WAREHOUSE and FER2013, of different variations in identity, emotions, and expressions. The mesh model would be able to seize the overall shape of the 3D face model.
- Smooth deformation is applied by estimating light conditions and reflectance on a Coarse Face(CF) model to achieve medium-scale features on the face.

The paper has been organized into various sections like related works, proposed method, experimental results, and conclusion.

II. RELATED WORKS

Initially,in the WELFake dataset [1] the position of major facial features such as the eyes, nose, and mouth, for example, are all comparable in human faces. From the standpoint of perception, it has been demonstrated that a face may be classified using only a few factors [1,2]. As the face space comprises very less dimensions, a constructive and effective parametric face representation could be obtained from a collection of sample faces. This leads to simplifying the reconstruction problem to parameter space searching. The 3DMM presented in [3] is

Doraikannan Sumathi is with the Department of Computer Science and Engineering, Vellore Institute of Technology, Andhra, India (e-mail: sumathi.research28@gmail.com)

Varanasi LVSKB Kasyap is with the Department of Computer Science and Engineering, Vellore Institute of Technology, Andhra, India (e-mail: varanasikasyap@gmail.com)

Mure Sai Jaideep Reddy is with the Department of Computer Science and Engineering, Vellore Institute of Technology, Andhra, India (e-mail: jaideepm05@gmail.com)

Varanasi Srinivasa Bhagavan is with the Department of Engineering Mathematics, KLEF, India (e-mail: drvsb002@gmail.com)

Thangamuthu Poongodi is with the School of Computer Science and Engineering, Galgotias University, India (e-mail: tpoongodi2730@gmail.com)

Thangamariappan Ganesh Kumar is with the School of Computer Science and Engineering, Galgotias University, India (e-mail: tganeshphd@yahoo.com).

Santhosh Kumar SVN is with the Department of Information Technology, Vellore Institute of Technology, Vellore 632014, India (e-mail: kumar.kumar34@gmail.com).

deployed in face processing tasks; reconstruction [3,4,5,6,7] and recognition [8,9]. These are well-known examples of such representations. Low-dimensional representations have also been employed in dynamic applications. Cao et al. [10] released FaceDBv2, a database containing the FG of 150 people of various ages and cultural origins. Similar to [11], our Face Modeling (FM) method uses a bi-linear face model to represent distinctiveness and expression aspects.

Shape-from-shading (SFS) is a computer vision algorithm that recovers 3D shapes from 2D image shading variations [12,13]. When the information regarding illumination, camera projection, and surface reflectance is provided, SFS algorithms can extract fine geometric aspects that may not be accessible using low-dimensional models. SFS, on the other hand, is a poor problem with ambiguous solutions [14]. Face reconstruction requires the application of past knowledge of facial geometry to create reliable results. Various researchers leveraged the symmetry of human faces to lessen the uncertainty of SFS results [15,16,17]. Another strategy is to address SFS using low-dimensional representation in facial space [18,19].

To find the posture of a 3D face, Li et al. [20] propounded a Face Reconstruction (3DFRC) technique based on 3D and coarse-fine matching. An adaptive method was used to produce the 3D model. The advantages of the method are partial occlusions and extreme situations was its durability. The model's performance is not promising when 2D and 3D facial landmarks are erroneously approximated for occlusion. Feng et al. [21] suggested a 3DFRC technique based on texture coordinates and UV position maps, and this technique is named a Position Map Regression Network (PRN). For regressing 3D face shapes from 2D pictures, Liu et al. [22] presented an encoder-decoder network. The joint loss was calculated based on 3DFRC and identification fault. However, the quality of face forms is found to have a great impact due to joint loss function. Using DCNNs and GANs, Gecer et al. [23] presented a 3DFRC. GAN was utilized for training the facial texture generator in UV space. A novel 3DMM fitting approach was developed on Generative Adversarial Network. Deng et al [24] introduced a reconstruction technique that works for especially weakly supervised learning. The losses at the perception and image levels were pooled. Large posture and occlusion invariant are two advantages of this method. During the prediction phase, however, the model's confidence is low on occlusion. Chen et al. [25] developed a self-supervised 3DMM trainable VGG encoder for 3D face reconstruction. A two-stage architecture was employed to regress 3DMM parameters to recreate facial details. Under normal occlusion, good-quality faces are created. UV space is used to record the face's details. On the other hand, the model fails to account for excessive occlusion, expression, and a huge position. The CelebA dataset [26] was utilized for training, while the LFW dataset [27,28,29] was used in conjunction with CelebA for testing. Ren et al. [28] created an encoder-decoder framework for 3D face point video deblurring. The distinctiveness knowledge and the structure of the face were predicted with the help of the reconstruction of the 3D face and the rendering branch [30]. Zhao et al. [31] worked on a method that can generate 3D face models using GANs. However, this approach is computationally intensive and less accurate to achieve facial textures. This approach in followed in addition with coaxial analysis by Patel el al. and Blanz et al. [32,33]. Mellouk et al. [34] presented a review on recognizing facial emotion and focused more about on accuracy of different approaches. Lin et al. [35] discussed about generating 3D model faces using graph convolutional networks, the results are relatively good but the illumination works are yet to be developed. Feng et al. and Khanzada et al. discussed about the metrics of measuring emotions of face in the papers [36,37]. There have been many works carried out to generate emotions on faces using classical deep learning methods [38,39,40,41]. However, it is a laborious and time-consuming task with less accurate results to generate emotions on the face. Deblurring of the face is also made possible. The limitations that exist in the existing face modeling works are enumerated as follows:

- Under a few circumstances, the texture quality of the generating face seems to be low.
- Facial alignment is not proper, which results in the generation of faces at different angles.
- Lighting on the face might lead to weird results.
- The usage of the joint loss function creates an impact on the quality of the face that is generated through the model
- A few methods are cost-effective. Preprocessing stages are found to be more expensive.
- Fail to construct the black-skinned faces

Considering all these limitations, the proposed work has been designed.

III. Proposed Methodology

The proposed work provides significance to constructing the 3D face models and emotional intelligence. The principle behind the design is the Face Vertex Land marking and Wulcheir distance. The following are the contributions towards achieving the objectives in this work are as follows:

- Facial features are extracted through the Deep Convolution Neural Network (DCNN).
- The proposed work combines the CoarseNet and FineNet to construct the 3D coarse face.
- The proposed DEO Model works based on plethora of 3D FM approaches i.e, DL, Epiploic Geometry, and the One-shot learning method.

The proposed methodology consists of four phases. Face Landmarks, construction of coarse images, transformation to Medium and the output phase. Height field face surface is computed by the shading deviations of the 2D input image and illumination parameter to enhance the face model. The final face model generated captures all the geometric details of the face, as shown in Fig. 1.

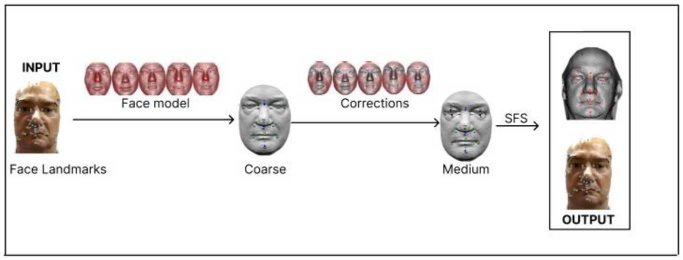

Fig. 1. Coarse-to-fine face Construction methodology

The reliable estimation of illustration parameters is achieved using the coarse method, and the robustness of expressions and the geometric details of the face model are computed with SFS. The end-to-end network proposed in this work connects the face construction problem and the other CNN applications like facial emotional recognition and facial recognition that allow further improved results with advanced CNN architectures. It also achieves the fast reconstruction of the face without using post-processing methods and external optimization. At the core, the idea is to divide the whole reconstruction into two modules, each composed of the neural network architecture. First, CoarseNet is introduced, a network to compute the face pose alignment and the geometric facial features. The rough facial geometric features are modelled with the help of 3DMM and this provides a dense network that can be recuperated by the proposed network. Next, FineNet is used to capture the fine information of the face. It functions on depth maps and is not restricted to the morphable model. To connect the CoarseNet output and FineNet a novel layer is taken that holds the backpropagation to jointly train both the networks. The proposed network gains state-of-the-art results and surpasses the existing SFS and example-based methods in terms of accuracy and fine detail construction evaluated on the public datasets. The next section describes the face modelling which provides the dynamic features for processing the face.

A. Face model (Low dimensional FM)

Human faces have universal characteristics such as eyes, nose, and mouth and are designated landmark vertex. The low-dimensional face space modelling is used for dynamic face processing and for transferring the facial performance between individuals. The low-dimensional representations are obtained from the example face shapes that could result in 3D face models. During the face reconstruction algorithm development, various face datasets are collected that are publicly available. The proposed CF modelling adopts the bilinear face model that encodes the identity and expressions of the face attributes similarly.

IV. Coarse Face Model Computing

A. Facial landmarks

The frontal facial pose is considered to mark the landmark vertex. For the non-frontal images, 2D landmarked face images do not respond with the vertices. Keeping the internal vertices unchanged, silhouette vertices are optimized according to the rotation matrix R. Original face mesh is pre-processed to obtain a dense horizontal line connecting mesh vertices and silhouette region from a rotated view. According to camera constraints π, R, t horizontal lines are projected onto an image plane, giving the estimated silhouette of the facial mesh projection. In silhouette landmark, a nearest analogous vertex is updated, as shown in Fig. 2.

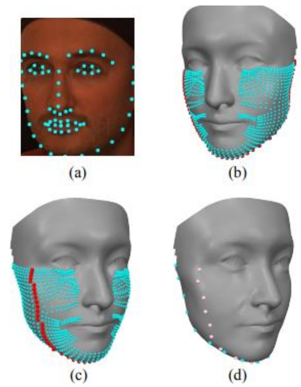

Fig. 2. Non-frontal face image labeled with 3D face silhouette landmarks updated (b) for better correspondence of detected silhouette landmarks.

Silhouette vertex is determined on a horizontal line, whose normal passes through the largest angle with a view of direction. The unit normal of the vertex is approximated as $\frac{R_v}{|R_v|^2}$ where v is the vertex coordinates, and the smallest silhouette vertex value is given as $|Z.\ \frac{R_v}{|R_v|^2}\ |2$ within the horizontal line, where Z= [0,0,1] T in the direction view. The silhouette improves the coarse model face image accuracy for non-frontal images. If there is no silhouette update, the resulting face models are erroneous. When the face is highly angular (more than 60◦), the coarse face model is less accurate unless the invisible facial landmarks are detected. This pipeline can also be combined with the pose-detection algorithms to achieve good results.

V. Medium Facial Feature Modelling

The CF model produces the face with the overall shape. However, the person-specific geometric features on the face are not captured due to the dataset constructional variance. The CF model is enhanced to get medium-scale features, improving the consistency between the shaded image and input image using the smooth deformation. In this process, lighting and albedo are also estimated, providing the prior knowledge to make SFS reconstruction further. The Single 2D images taken are converted into grayscale images for easy computation and efficiency.

A. Albedo and lighting estimation

Shading for the face mesh is computed using the lighting and surface reflectance information. The grayscale level at pixel (x,y) is estimated by second-order spherical harmonics assuming Lambertian reflectance:

$$S_{x,y} = r_{x,y} \ * \ max \ (\varepsilon^T \ SH(A_{x,y}),0) \tag{1}$$

Where $r_{x,y}$ is the albedo at the pixel; ε^T is the harmonic coefficient vector; $A_{x,y}$ is the corresponding mesh regular, calculated as

$$A_{x,y} = \frac{(V_2^{x,y} - V_1^{x,y}) * (V_3^{x,y} - V_1^{x,y})}{||(V_2^{x,y} - V_1^{x,y}) * (V_3^{x,y} - V_1^{x,y})||^2} \tag{2}$$

Where $V_1^{x,y}, V_2^{x,y}, V_3^{x,y}$ are the mesh triangle vertex coordinates at the pixel (x,y); SH is the second-order spherical harmonic vector:

$$SH(n) = [1, n_x, n_{y,} n_z, n_x n_{y,} n_x n_{z,} n_y n_{z,} n_x^2 - n_y^2, 3n_z^2 - 1]^T \tag{3}$$

The Principal Component Analysis model is used here to parametrize the surface reflectance given as:

$$r_{x,y} = [\varphi + \sum_{i=1}^{A} w_r^l \varphi_l] * b_{x,y} \tag{4}$$

Where $[V_{x,y}^1, V_{x,y}^2, V_{x,y}^3$ are the coordinates of the barycentric triangle $r_{x,y}$.

B. Enhancing facial details

Once the estimation of the lighting and reflection is done, the CF model is enhanced to scale down the input image and mesh shading discrepancy. The low dimensional smooth face is constructed using the norm of the Laplacian graph concerning the mesh. Since most frontal face variations are specific to the person, local regions of mesh are considered to perform Laplacian Eigen analysis separately. Medium face modeling improves the facial expressions, lighting, and albedo conditions more accurately. The comparison of the overall face to the coarse face model generated is calculated using the basis vectors, similar to finding the corresponding Fourier analysis of the Laplacian matrix. The facial features such as dimples and laugh outlines are not reconstructed using the CF model. Identity parameters of the CF model are altered and further constructed using the medium facial features resulting in the overall shape of the face.

VI. Detailed face Modelling

A. Normal map optimization

The computed normal is enforced to obtain accurate geometric details and desirable properties on the face. Using the lighting and albedo conditions of the face obtained, pixel intensity values are calculated from the normal map. However, inaccuracy is obtained while calculating the second-order spherical harmonics in complex lighting and albedo conditions. So, intensity gradients are minimized in the shaded image from the normal given as:

$$\boldsymbol{I_{grad}} = \sum ||[\begin{matrix} \boldsymbol{S_{x+1,y}^1} - \boldsymbol{S_{x,y}^1} \\ \boldsymbol{S_{x,y+1}^1} \boldsymbol{S_{xy}^1} \end{matrix}] - [\begin{matrix} \boldsymbol{I_{x+1,y}^1} - \boldsymbol{I_{x,y}^1} \\ \boldsymbol{I_{x,y+1}^1} \boldsymbol{I_{xy}^1} \end{matrix}]||^2 \tag{5}$$

Since the problem is under-constrained, for regularizing the intensity gradient additional, regularizing terms are added to the normal map.

$$\boldsymbol{I_{reg}} = ||\boldsymbol{n_{x,y}^1} - \boldsymbol{n_{x,y}^0}||_2^2 \tag{6}$$

B. Combined network training

CoarseNet is connected to the FineNet using the rendering layer, which transforms the 3DMM representation of the depth map. The gradient is propagated through the network to train both the CoarseNet and FineNet simultaneously, as shown in Fig. 3. The criterion is the mean squared error between the CoarseNet and the FineNet weighted and passed to tune it.

Fig. 3. The gradient from both loss criteria and propagated back to CoarseNet (Criterion flow).

VII. Experiments and Results

This section presents the experimental results of the implemented method and compares the proposed method with existing methods. The Bosphorus database is used to verify the effectiveness of the technique. This database comprises the structured-light 3D face clouds of 105 people and their single-angle 2D face images. From this database, 55 images are chosen with low gaussian for testing. The face is aligned and cropped at a radius of 90mm, taking the nose as the center and computing the 3D Root Mean Square Error (3DRMSE). The proposed method is enacted on a PC with Intel Core i7 with 2.5 GHz CPU and 8 GB RAM.

A. Frontal face results

The mean and standard variation of the 3DRMSE is 1.87 ± 0.25, and 1.92 ± 0.19. This proves that the result's mean error is consistently lower than the existing methods [3,8,9,18,23]. The comparison results of the existing methods are also taken, and the results are illustrated in Fig. 4.

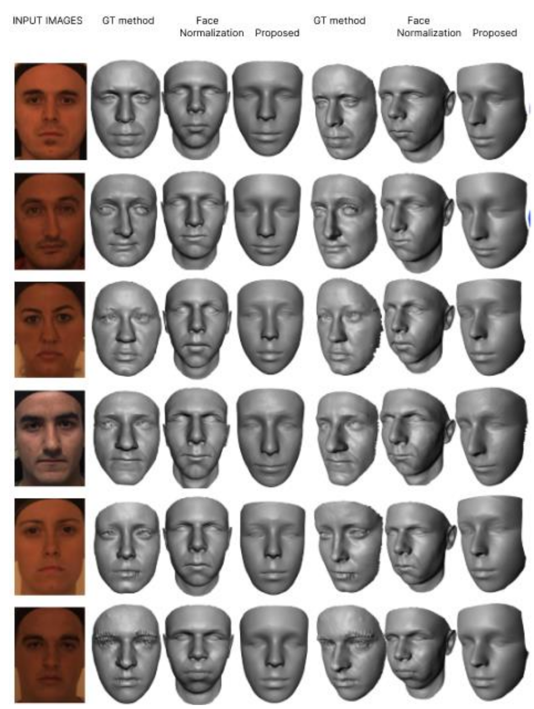

Fig. 4. Facial reconstructions of GT method, face normalization, proposed method.

B. *Non-frontal face results*

Non-frontal images are also used to test the model. Using landmark detection methods, facial images with the pose can be reconstructed. The obtained results look constant for various poses and expressions, and the errors are found to be small. Each method provided has an estimated depth image and a binary representation of the pixels, and the proposed method has a lower depth error than the existing methods. For the qualitative analysis of the results, the method is evaluated on the FRG challenge dataset, with a mean and standard deviation of 3DRMSE is 1.47 ± 0.15, and 1.64 ± 0.09. The results are shown in Fig. 5.

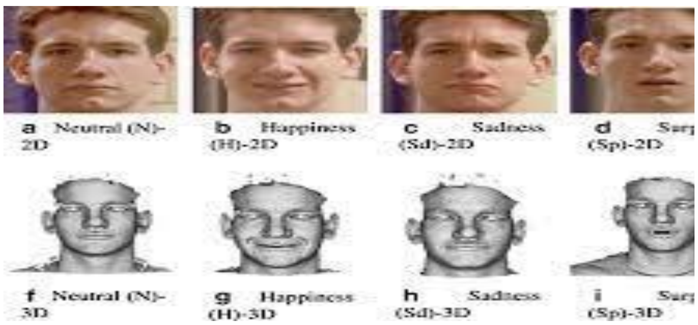

Fig. 5. Non-frontal face reconstructions of one persons with expressions and poses

In this work, the proposed methods resolve all the problems stated above. It generates the High textured 3D model with fine details using the CoarseNet and FineNet. The processing time and construction time are faster than the existing methods. The proposed method is found to be robust to partial occlusions and extreme poses by generating the face clouds. The advantage of the 3D component is it requires less modeling and is easy to handle. The proposed method uses the dual architectural benefit of both CoarseNet and FineNet, thereby decreasing face construction errors. The computational time required by the proposed method is 3.2 seconds, 2.5 seconds, 4.6 seconds, and 3.1 seconds on FaceDBv2, Bosophorous, UHDB31, BU-3DFE, and FRGC2 databases, respectively. The construction error of the faces is shown in Fig. 6.

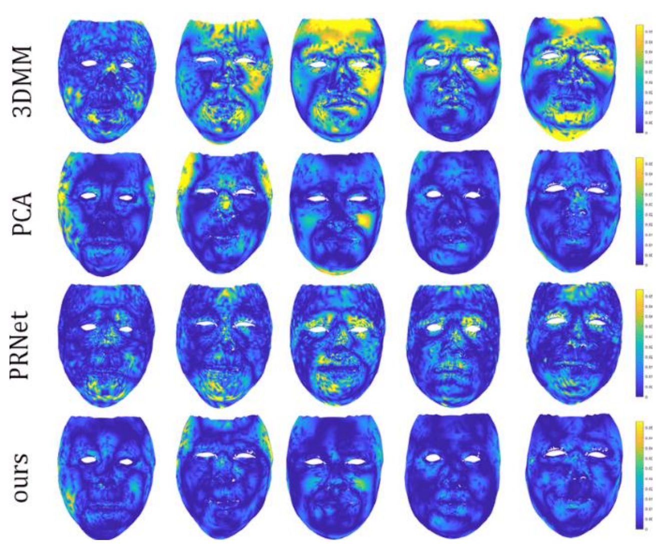

Fig. 6. The construction error heatmaps on FRGC2 database of different methods. This illustrates the proposed method has a negligible error than the existing methods.

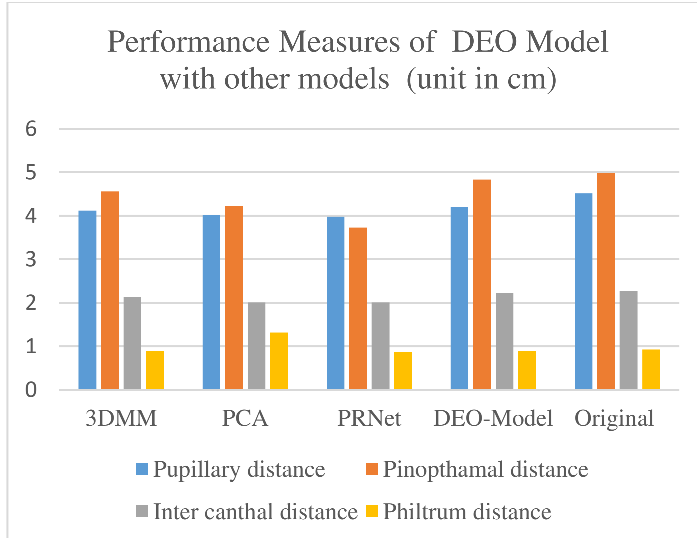

Fig. 7. Performance comparison

The performance metrics of various models have been provided in the Fig 7. The comparison of the cumulative distribution of the RMSE error of different models on various databases is given in Table 1.

Table I. Performance measures of the proposed algorithm

	[1]	[2]	[5]	Ours
FaceDBv2	N/A	3.24 ±0.4	2.72 ±0.3	**1.34 ±0.3**
Bosophorous	3.78 ±0.9	2.79 ±0.6	2.72 ±0.3	**1.76 ±0.6**
BU-3DFE	2.89 ±0.5	3.57 ±0.1	4.59 ±0.2	**2.56 ±0.5**
FRGC2	4.32 ±0.3	3.95 ±0.1	3.62 ±0.7	**3.21 ±0.4**

Fig. 8 depicts the RMSE comparison of various models and the proposed model outperforms the other model.

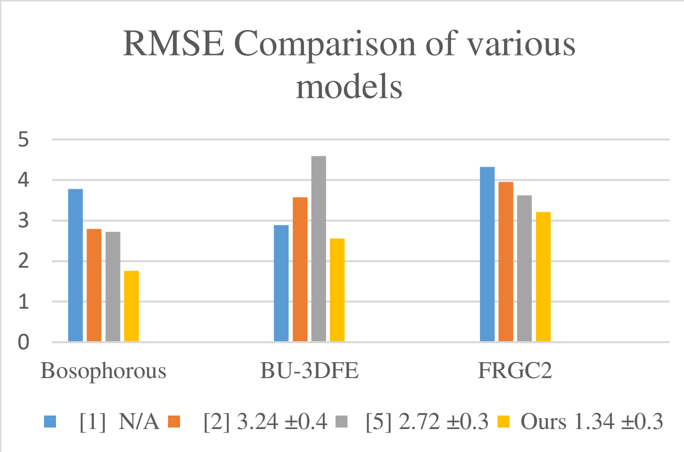

Fig. 8. Comparison of RMSE for various models.

C. *Similarity measures*

The purpose of the 3D face similarity measure is to compare computer-generated 3D facial models and determine how similar they are. Metrics based on various distances such as geodesic distance, Hausdorff distance Frechet distance and Euclidean distance are used. Among these distances, geodesic distance plays a significant role in generating the expression-invariant representation of faces. Additionally, it is also found to be not consideration with the changes in the facial expression.

D. *Geodesic curvature similarity measure (GCS measure)*

Using the geodesic curve, we can measure the local minimum curve in the region of two planar points of the facial mesh. This method is efficient evaluation metric to compare the real face and 3D model. This method is intrinsic and is insignificant to the curved modelling of the 3D mesh model. The geometric transformation of the 2D image to the 3D model is the analyzed to find the similarity score between both. Bronstein et al. [42] worked on Euclidean distance and geodesic distance and concluded that the geodesic distance is more efficient way to find the similarity of the non-liner models. This analysis is proved by using the 133 face markers inflated on the image and computing the geodesic distances between the markers. In this study, under the small facial deformations and similar reformations, the facial surfaces in the generated models are geometric. This method is used in evaluating the similarity of the faces because it uses all the geodesic properties to generate the expression-invariant facial models.

Geodesics of the model can be computed by backtracking the similarity of the faces. The primary point on the facial model can using the shortest geodesic distance between the facial markers used for the analysis. The other points of the geodesics of the traveling the adjacent and adjoint facial positions in return. The extracted geodesic regions of the facial model along with the angular change α in the facial model as given in equation (7).

$$J = Upper_{\alpha} * ground_{\alpha} \qquad \alpha \in P$$

$$A = \left\{ 0, \frac{2\pi}{k}, \frac{4\pi}{k} \ldots\ldots, \frac{2\pi(l-1)}{k} \right\} \qquad (7)$$

In (7), G refers to the geodesics set of 3D facial model and P denotes the angular positions corresponding to the geodesics model. In Fig. 9., 60n different positions of the geodesics from the 3D facial model are given.

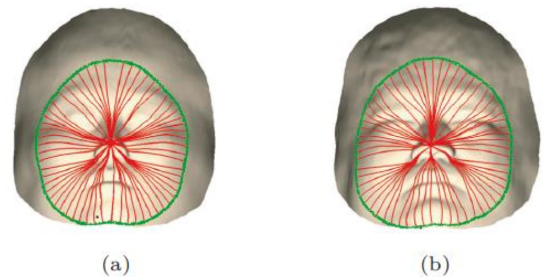

Fig. 9. Geodesic representation of the 3D facial models

E. *CNN based module for similarity measure*

To produce a quantitative evaluation of similarity between the 2D face image and the generated 3D facial model, a deep convolution neural network is used. This framework is used along with the Siamese network to compare the 3D facial mesh to a 2D image. This network is trained using the L2 distancing techniques in the embedding space of the 3D facial mesh plane. This is advantageous because of the computational intensity of the model. The proposed network is optimized using the loss function as shown in (8).

$$L = (1 - \tau)D(I1, I2)^2 + \tau * (0, k - D(I1, I2)^2) \qquad (8)$$

The (I1, I2) are the input images of the model, D(*) is the difference between the embeddings of the input image and facial model. The results of the proposed metric are given in Fig. 10.

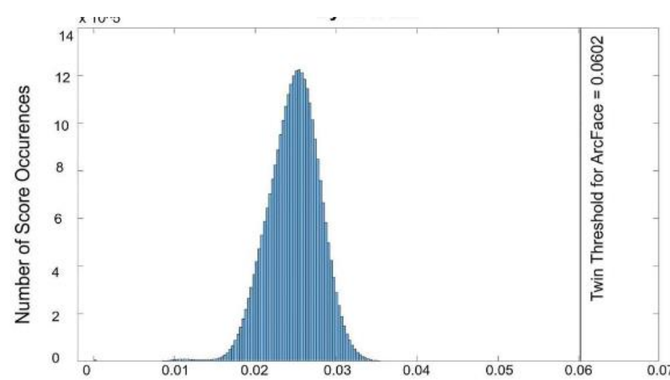

Fig 10. The score occurrences in the faces and the comparison of the facial expressions

F. *Face alignment similarity measure*

In this section of evaluation, an unsupervised approach for sequential assessment performance of the 3D modelling algorithms is shown. The 3D landmarks are extracted from the 3D facial data to approximate the transformations from the facial contour maps to the 3D facial mesh model. The discrepancy between the generated 3D facial models and the extracted facial landmarks are compared. Then the confidence score (C_S) is calculated. Based on this score, we can effectively analyze the localization of the 3D facial model. This technique enables to find the overall efficiency and performance of the proposed framework. The statistical landmark modelling is used in this study to compare and find similarity score. To achieve this, we used another dataset $D_{additional}$ is used that contains N face images with characteristics: prompted poses, expressions, both speaking and neutral faces, but extrinsic sourcing of perturbation such as presence of object that causes illuminations and occlusions. The extracted 3D facial landmarks from these datasets using a 3DFA algorithm robustly estimates the rigid transformation between the landmark-l dataset and the and the neural front load dataset as show in (9).

$$y_{n,l}^{alg} = Sample_l^{alg} R_l^{alg} x_{n,l}^{alg} + t_l^{alg} \tag{9}$$

The two different assertions here mean the posterior and geometric frontiers of the face scale of 0 to 1 for 3DFA framework given in equations 10 and 11. The correlation of distinct threshold values of calculated posterior and geometric functions are plotted in fig 11 and obtained results are stated in Table 2.

$$P_n^{gum} = \frac{\sum_{l=1}^{L} u_{n,l}^{p} x_{n,l}^{3DFA}}{\sum_{l=1}^{L} u_{n,l}^{p}} \tag{10}$$

$$C_n^{gum} = \frac{\sum_{l=1}^{L} u_{n,l}^{p} (x_{n,l}^{3DFA} - x_{m,l}^{3DFA}) * (y_{n,l}^{3DFA} - y_{m,l}^{3DFA})}{\sum_{l=1}^{L} u_{n,l}^{p}} \tag{11}$$

Table II. Comparison of 3DFA models with different training datasets

Data Set	Statistical facial model trained on $D_{additional}$ datasets						
	Face RD model [7]	FaceNet [9]	3DFA [11]	CartoonNet [19]	VR framewor [28]	GTH model [32]	DEO Model [41]
TD – 1	0.83	0.76	0.89	0.87	0.85	0.78	**0.95**
TD – 2	0.92	0.89	0.91	0.90	0.86	0.87	**0.94**
TD – 3	0.96	0.97	0.97	0.95	0.93	0.96	**0.98**
TD – 4	0.10	0.06	0.20	0.05	0.02	0.20	**0.35**
TD – 5	0.21	0.04	0.10	0.07	0.21	0.29	**0.34**
TD – 6	0.73	0.58	0.67	0.55	0.67	0.71	**0.81**
TD – 7	0.68	0.14	0.45	0.23	0.82	0.71	**0.88**
TD – 8	0.87	0.67	0.77	0.78	0.82	0.68	**0.92**

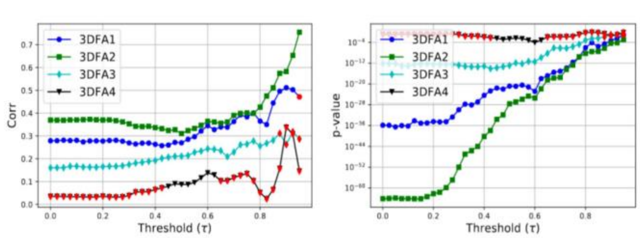

Fig 11. Comparison of threshold values of the DEO model with existing models

From the Fig 11 the comparative measures of the threshold values of the DEO model with the earlier works shows that the DEO model outperforms the other models.

VIII. Conclusion

This work proposes a coarse-to-fine method to construct a sharp and fine focused 3D face model from one image. This method adopts the bilinear facial approach along the corrective deformation field to construct a fine detailed facial model. The experiments show that the model accurately constructs the 3D facial shapes from images with different facial expressions, wrinkles, person-specific details, and poses. This proposed solution has dual benefits of low-dimensional techniques and SFS techniques with error-free facial construction. Since the albedo and lighting values are computed using pixels, PCA is used to avoid the inherent ambiguity generated by the noise of the input images. This is sufficient to achieve accuracy in the facial geometric details. The limitation of the method is the performance of the model depends on the overall shape of the input image and the angle of the facial pose. If the facial pose is more than 90°, then the method's performance might be affected. The potential future work is to extend the 3D model construction of the human body.

Declarations

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest

The authors declare no potential conflict of interests

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to privacy access rights but are available from the corresponding author on reasonable request.

Compliance with Ethical Standards

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

References

[1] Ren W, Yang J, Deng S, et al (2019) Face video deblurring using 3D facial priors. In: Proceedings IEEE int conf comput vis 2019-Octob:9387–9396. https://doi.org/10.1109/ICCV. 2019.00948

[2] Li K, Yang J, Jiao N, et al (2020) Adaptive 3D face reconstruction from a single image. 1–11. arXiv preprint arXiv:2007.03979

[3] Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. 2021. Learning an animatable detailed 3D face model from in-the-wild images. ACM Trans. Graph.40, 4, Article 88 (August 2021), 13 pages. DOI:https://doi.org/10.1145/3450626.3459936

[4] S. Romdhani and T. Vetter, "Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., vol. 2, Jun. 2005, pp. 986–993.

[5] M. Keller, R. Knothe, and T. Vetter, “3D reconstruction of human faces from occluding contours,” in Computer Vision/Computer Graphics Collaboration Techniques. 2007, pp. 261–273.

[6] A. Bas, W. A. Smith, T. Bolkart, and S. Wuhrer, “Fitting a 3D morphable model to edges: A comparison between hard and soft correspondences,” in Proc. Asian Conf. Comput. Vis. Taipei, Taiwan: Springer, 2016, pp. 377–391.

[7] P. Huber et al., “A multiresolution 3D morphable face model and fitting framework,” in Proc. 11th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl., 2016, pp. 79–86.

[8] Jiang, Luo & Zhang, Juyong & Deng, Bailin & Li, Hao & Liu, Ligang. (2017). 3D Face Reconstruction With Geometry Details From a Single Image. IEEE Transactions on Image Processing. PP. 10.1109/TIP.2018.2845697.

[9] V. Blanz and T. Vetter, “Face recognition based on fitting a 3D morphable model,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[10] 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.

[11] D. Vlasic, M. Brand, H. Pfister, and J. Popovi´c, “Face transfer with multilinear models,” ACM Trans. Graph., vol. 24, no. 3, pp. 426–433, 2005

[12] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, “Shape-from-shading: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 8, pp. 690–706, Aug. 1999.

[13] J.-D. Durou, M. Falcone, and M. Sagona, “Numerical methods for shapefrom-shading: A new survey with benchmarks,” Comput. Vis. Image Understand., vol. 109, no. 1, pp. 22–43, Jan. 2008.

[14] E. Prados and O. Faugeras, “Shape from shading,” in Handbook of Mathematical Models in Computer Vision, N. Paragios, Y. Chen, and O. D. Faugeras, Eds. New York, NY, USA: Springer, 2006, pp. 375–388.

[15] I. Shimshoni, Y. Moses, and M. Lindenbaum, “Shape reconstruction of 3D bilaterally symmetric surfaces,” Int. J. Comput. Vis., vol. 39, no. 2, pp. 97–110, Sep. 2000.

[16] M. Hernandez, T. Hassner, J. Choi, and G. Medioni, “Accurate 3d face reconstruction via prior constrained structure from motion,” Computers & Graphics, 2017.

[17] W. Y. Zhao and R. Chellappa, “Symmetric shape-from-shading using self-ratio image,” Int. J. Comput. Vis., vol. 45, no. 1, pp. 55–75, Oct. 2001.

[18] J. J. Atick, P. A. Griffin, and A. N. Redlich, “Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images,” Neural Comput., vol. 8, no. 6, pp. 1321–1340, Aug. 1996.

[19] R. Dovgard and R. Basri, “Statistical symmetric shape from shading for 3D structure recovery of faces,” in Computer Vision—ECCV, T. Pajdla and J. Matas, Eds. Prague, Czech Republic: Springer, 2004, pp. 99–113.

[20] Li K, Yang J, Jiao N, et al (2020) Adaptive 3D face reconstruction from a single image. 1–11. arXiv preprint arXiv:2007.03979

[21] Feng Y, Wu F, Shao X, et al (2018) Joint 3d face reconstruction and dense alignment with position map regression network. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11218 LNCS:557–574. https://doi.org/10. 1007/978-3-030-01264-9_33

[22] Liu F, Zhu R, Zeng D et al (2018) Disentangling features in 3d face shapes for joint face reconstruction and recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https:// doi.org/10.1109/CVPR.2018.00547

[23] Gecer B, Ploumpis S, Kotsia I, Zafeiriou S (2019) Ganft: Generative adversarial network fitting for high fidelity 3D face reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition:1155–1164. https://doi. org/10.1109/CVPR.2019.00125.

[24] Deng Y, Yang J, Xu S, et al (2019) Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. IEEE Comput Soc Conf Comput Vis Pattern Recognit Work 2019-June:285–295. https://doi.org/10.1109/CVPRW. 2019.00038

[25] Chen Y, Wu F, Wang Z et al (2019) Self-supervised Learning of Detailed 3D Face Reconstruction. IEEE Trans Image Process 29:8696–8705

[26] Large-scale CelebFaces Attributes (CelebA) Dataset. http:// mmlab.ie.cuhk.edu.hk/projects/CelebA.html. Accessed 13 Oct 2020

[27] Labelled Faces in the Wild (LFW) Dataset | Kaggle. https://www. kaggle.com/jessicali9530/lfw-dataset. Accessed 13 Oct 2020

[28] Ren W, Yang J, Deng S, et al (2019) Face video deblurring using 3D facial priors. In: Proceedings IEEE int conf comput vis 2019-Octob:9387–9396. https://doi.org/10.1109/ICCV. 2019.00948

[29] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, “High-fidelity pose and expression normalization for face recognition in the wild,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 787–796.

[30] Richardson, E., Sela, M., Or-El, R., & Kimmel, R. (2017). Learning Detailed Face Reconstruction from a Single Image. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5553-5562.

[31] Zhao, Dapeng & Cai, Jinkang & Qi, Yue. (2022). Convincing 3D Face Reconstruction from a Single Color Image under Occluded Scenes. Electronics. 11. 543. 10.3390/electronics11040543.

[32] N. Patel and M. Zaveri, "3D Facial model construction and expressions synthesis from a single frontal face image," 2010 International Conference on Computer and Communication Technology (ICCCT), 2010, pp. 652-657, doi: 10.1109/ICCCT.2010.5640447.

[33] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3D faces,” in Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH), 1999, pp. 187–194

[34] Wafa Mellouk, Wahida Handouzi, Facial emotion recognition using deep learning: review and insights, Procedia Computer Science, Volume 175, 2020, Pages 689-694, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.07.101.

[35] Lin, Jiangke & Yuan, Yi & Shao, Tianjia & Zhou, Kun. (2020). Towards High-Fidelity 3D Face Reconstruction From In-the-Wild Images Using Graph Convolutional Networks. 5890-5899. 10.1109/CVPR42600.2020.00593.

[36] Y. Lv, Z. Feng and C. Xu, "Facial expression recognition via deep learning," 2014 International Conference on Smart Computing, 2014, pp. 303-308, doi: 10.1109/SMARTCOMP.2014.7043872.

[37] Khanzada, Amil & Bai, Charles & Celepcikay, Ferhat. (2020). Facial Expression Recognition with Deep Learning.

[38] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

[39] Fu, Haibin & Bian, Shaojun & Chaudhry, Ehtzaz & Iglesias, Andres & You, Lihua & Zhang, Jian. (2021). State-of-the-Art in 3D Face Reconstruction from a Single RGB Image. 10.1007/978-3-030-77977-1_3.

[40] Hu, Weirong & Li, Yu & Wang, Xuesong. (2021). Research on Current Situation of 3D face reconstruction based on 3D Morphable Models. Journal of Physics: Conference Series. 1966. 012017. 10.1088/1742-6596/1966/1/012017.

[41] Dou, Pengfei, S. Shah and I. Kakadiaris. “End-to-End 3D Face Reconstruction with Deep Neural Networks.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 1503-1512

[42] Bronstein A M, Bronstein M M, Kimmel R. Threedimensional face recognition. International Journal of Computer Vision, 2005, 64(1): 5-30.