SL2E-AFRE : Personalized 3D face reconstruction using autoencoder with simultaneous subspace learning and landmark estimation

P. R. Suganya Devi[1] [ORCID] · R. Baskaran[1]

Abstract

3D face reconstruction from single face image has received much attention in the past decade, as it has been used widely in many applications in the field of computer vision. Despite more accurate solutions by 3D scanners and several commercial systems, they have drawbacks such as the need for manual initialization, time and economy constraints. In this paper, a novel framework for 3D face reconstruction is presented. Firstly, landmarks are localized on the database faces with the proposed landmark-mapping strategy employing a model template. Then, an autoencoder assisted by the proposed energy function to simultaneously learn the facial patch subspace and the keypoints positions is employed to predict the landmarks. Finally, an unique 3D reconstruction is obtained with the proposed predicted landmark based deformation. Meta-parameters are incorporated into the energy function during the training phase to enhance the performance of the autoencoder network in reconstructing the face model. The experiments are carried out on two databases namely the USF Human ID 3-D Database and the Bosphorus 3D face database. The experimental results show that the Autoencoder based Face REconstruction with Simultaneous patch Learning and Landmark Estimation method (SL2E-AFRE) is efficient and the performance of the same is significantly upgraded in each iteration.

Keywords 3D Face reconstruction · Autoencoder · Landmark estimation · Shape deformation

1 Introduction

Three-dimensional knowledge about a human face is very much beneficial for computer vision and computer animation fields. 3D Face Reconstruction can be defined as promoting a 2D face image into a 3D geometry. Since a 3D face model provides a mathematical representation of the face surface in three dimension, which is consistent despite change in the pose, illumination and expression, 3D face reconstruction is therefore essential for real-time applications in face recognition, plastic surgery, facial animation [10, 29], entertainment [9] and computer games [38, 39], 3D rendering of comic cartoon characters [26, 27] etc. In the media and entertainment industry(M&E),

it is contemporary to use 3D rendering of human faces in various segments such as cinema, advertising and gaming. In the medical field, 3D printing exhibits abundant applications such as 3D modeling of human organs from computed tomography(CT) data [43]. 3D reconstruction is essential to generate 3D-printed model of the anatomical structure for pre-surgical preparation and also to carry out cosmetic surgery. In criminal investigation [14, 32, 41], facial recognition is of predominant importance and 3D technology is used for precise recognition.

A 3D face model can be represented by polygonal mesh or B-spline. The ambiguity feature of the human face limits the accuracy of 3D face modeling. 3D point cloud obtained by the 3D scanners could be the best input for generating 3D face model. But it is very expensive. So, researchers started focusing on 3D face reconstruction from 2D images captured by the camera. Also, acquiring multiple views of a face is not feasible in all the real time applications. So, using a single face image as input for 3D modeling gains more importance. Irrespective of the methods used to capture the input, the result will be greatly affected by the person's pose and illumination. Several 3D face reconstruction methods have been presented in the state-of-the art. However, dealing

✉ P. R. Suganya Devi
  suganya.rvks5@gmail.com

  R. Baskaran
  baaski@annauniv.edu

1 Department of Computer Science and Engineering, College
   of Engineering, Guindy, Anna University, Chennai, India

with pose and illumination variation is still a significant problem. This affects the semantic details of the resultant 3D model.

Many datasets include images and 3D shapes annotated with the primary (eyes,mouth,nose) and secondary(facial outline,chin,hair) face features, which is not sufficient for detailed 3D face geometry reconstruction. Therefore, to handle cheek and contour information, the intermediate serving keypoints(ISK) are introduced in this work. And, in some datasets, though such uncalibrated images and models are there, the keypoints are not available, which is crucial for learning based 3D modeling. Therefore, in order to prepare the publicly available 3D face databases without required keypoints, to be useful for research purpose, a Landmark-mapping strategy is proposed in this work, to localize the landmarks on the faces in the database, adopting a model template consisting of required keypoints. It needs no manual intervention. The proposed landmark-mapping strategy can map a set of landmarks from one template model to any arbitrary model.

Most of commercial systems have restrictions in modeling 3D shape such as the requirement of frontal and profile views, manual localization of feature points. In the proposed work, the problem of 3D face modeling from a frontal face image with automated feature localization using cluster analysis [16, 40, 42] is particularly focused. Instead of generating 3D face using the generic model, the SL2E-AFRE follows predicted landmark based reference model deformation approach resulting in an unique 3D shape model divergent from the generic model. The literature which has given high accurate 3D models have used the deep learning for both 2D as well as 3D subspace learning purpose. The proposed approach is confined to learn the 2D subspace only using autoencoder, which is basically a dimensionality reduction technique. Since the regression based landmark estimation [44] is sensitive to the face appearance inside the bounding box [45], clustering based landmark estimation is proposed in this work.

1.1 Contributions

As key contribution, this work presents a new energy function to be optimized within the autoencoder architectures. The energy function simultaneously learns the subspace and landmarks positions. The overall contribution of this paper are as follows: In this work, a three-step 3D face reconstruction approach is presented which includes adapting the database to the proposed 3D reconstruction framework, landmarks estimation and shape deformation.

1. In the first step, a set of Intermediate Serving Keypoints(ISK) is computed on the model template and an enhanced keypoint set is formed using ISK along with the base keypoints to derive more accurate facial shape. Then, the proposed landmark mapping strategy is applied to adapt the 3D face database to the proposed 3D face reconstruction approach by mapping the landmarks from the model template to the database, which does not require any manual intervention.

2. In the second step, a novel architecture based on auto encoders for 3D face landmarks prediction is presented, where learning the patch subspace and Landmark positions are carried out simultaneously with the help of the proposed energy function. Only 2D images are learned with the deep network, which can dramatically reduce computation.

3. In the third step, a deformation method based on predicted landmarks is proposed to obtain a personalized unique 3D face shape, where instead of using a generic face, a reference model to be deformed is selected from the training examples based on the predicted landmarks. A model is selected by performing interpolation followed by finding distance between predicted landmarks and landmarks of training examples.

4. The proposed system is tested with two databases, namely The Bosphorus 3D Face Database [31] and the USF Human-ID 3D Face database [6].

The performance of proposed approach is comparable to other state-of-the-art works. However, the proposed approach is far faster than the existing methods as it does not employ complex architectures.

This paper is formulated as follows. Section 2 describes related works. A new methodology for 3D face reconstruction is presented in Section 3. Section 4 analyses the experimental results obtained. Section 5 concludes the paper.

2 Related work

Existing 3D face reconstruction approaches can be classified into three categories: 3D Morphable Model(3DMM), Shape-from-Shading(SFS), Learning-based Methods. Shape from Shading methods determine the shape using shading information(brightness variation from one pixel to another) inherited from the image [18]. The main drawback of the SFS method is that its presumption of lambertian surface and a single point light source at infinity which led to unrealistic 3D face reconstruction [8]. In [23], Kemelmacher et al. exploit the similarity of faces and combines the input image's shading information with that of the generic model. To overcome the problem arose by the assumptions, the researchers have proposed various approaches [35]. However, SFS does not provide an unique solution owing to the complex albedo variations in the face. Since SFS captures the fine-scale facial geometric details, it is being used as the tune-up phase in some approaches [21].

In 3DMM approach, a deformable model is generated based on a linear combination of example 3D models [6]. Then the parameters are estimated by optimizing the cost function, which is the difference between the 2D projection of the model and the input image. The 3D model can be generated using the estimated parameters. Because of using the intensity information while fitting the model, the process becomes time-consuming and most probably trapped into local minima, which results in unrealistic 3D geometry. To overcome the local minima problem, researchers started using the facial landmarks in the deformable model to fit those in the input 2D image [20]. These approaches require human intervention to locate the landmarks and the estimated parameters are less accurate. Zhou et al. [49] introduces shape-space model, where each basis shape can be rotated and so it is robust to arbitrary initialization. Baumberger et al. [4] used silhouette information to deform the generic model. Dou et al. [12] proposed a deep neural network based End-to-End 3D face reconstruction and generate 3D model from unconditional inputs. Richardson et al. [30] generates a coarse face model using CNN, employs another network to refine the coarse model and uses a rendering layer to connect the networks. Tran et al. [36] employs bump mapping method, in which global shape is computed before local features estimation. Ding et al. [11] presented a local linear fitting (LLF) based 3D face reconstruction with sparse key points. Luo et al. [25] used ICF (Iterative Closest Point) algorithm, where each vertex is provided with an optimal weight for aligning the 3DMM to the given depth image.

Learning based methods used to learn the 2D images subspace and the corresponding 3D models subspace. Then the mapping can be computed between these two subspaces to generate the 3D face model for the input face image [47]. Han et al. [15] proposed an approach based on cascaded regression, where a shape incremental feature was used, which exploits the information from 2D face and the current estimated model at that stage. It could not handle pose variations. In [33] and [24] coupled radial basis function is used to get the intermediate face. The model is optimized through landmarks using a coupled dictionary, which relates 3D face model and 3D landmarks, and another coupled dictionary, which relates 2D and 3D landmarks(to obtain z-coordinates). Sun et al. [34] proposed a coupled statistical model which incorporates both face image and depth map of the face. In this approach, a new database is generated by illuminating the training dataset with an illumination parameter estimated from the average face model and the input image. This method is robust to different light conditions. Zhang et al. [48] used the Stacked Contractive Autoencoder (SCAE), which learns nonlinear image subspace and corresponding 3D model subspace. Then a one-layer neural network is used to compute the mapping between the subspaces. The drawback of this approach is its computation complexity. Similarly, Arslan and Seke [2] used conditional generative adversarial networks(CGAN) for computing depth. Jackson et al. [19] and Feng et al. [13] learn the mapping from 2D to 3D coordinates using CNN(Convolutional Neural network). Jackson et al. [19] develops the volumetric representation, which however does not consider the semantic significance of the points but the method proposed by Feng et al. [13] does. Tran et al. [37] used deep neural networks (DNN), where in-the-wild images are used for training.

3 Methodology

The proposed approach is succinctly illustrated in Fig. 1. It includes three components: 1) landmark mapping across 3D databases; 2) patch subspace learning and facial landmark estimation with autoencoder; 3) predicted landmark based deformation.

Initially, we need to find N 3D facial landmarks $L = \{l_1, l_2, ..., l_N\}$, in the normalized 3D mean frontal face geometry. In this work, to localize the landmarks in the 3D face model, a landmark mapping approach across database is presented. Then, the reconstruction problem is solved by simultaneous clustering and relevant keypoints estimation procedures by incorporating the proposed energy function within the autoencoder network. The Energy function to be minimized to find the landmarks is as follows:

$$E = AE_{loss} + \alpha.landmark_{loss} - \beta.cluster_{loss} \tag{1}$$

where AE_{loss} ensures the consistent result though the code layer data of autoencoder is used. $landmark_{loss}$ is the error in the landmarks predicted. Since the code layer is the base for clustering, which is the base for landmark prediction and the input is not represented precisely in the code layer, $cluster_{loss}$ takes a part in this energy function.

Autoencoder network is used with the intention of compressing the dimension, since it can learn the non-linear mapping in an unsupervised way effectively. Autoencoder maps the high dimensional data to low dimensional space. Existing deep learning based approaches [12, 19] obtain non-linear subspaces of both 2D and 3D samples, entailing computationally intensive methods. In this proposed approach, only 2D subspace of facial patches is learned with autoencoder network. Then 3D geometry of different patches of a face is generated, each from different individuals, by performing cluster analysis and 3D landmark estimation simultaneously. Finally, a reference model chosen from the training dataset based on the prediction of landmarks, is deformed using the estimated keypoints by the laplacian deformation technique. In this paper, the set of 3D face landmarks is represented as $L = \{l_1, l_2, ..., l_N\}$, and each 2D face image is
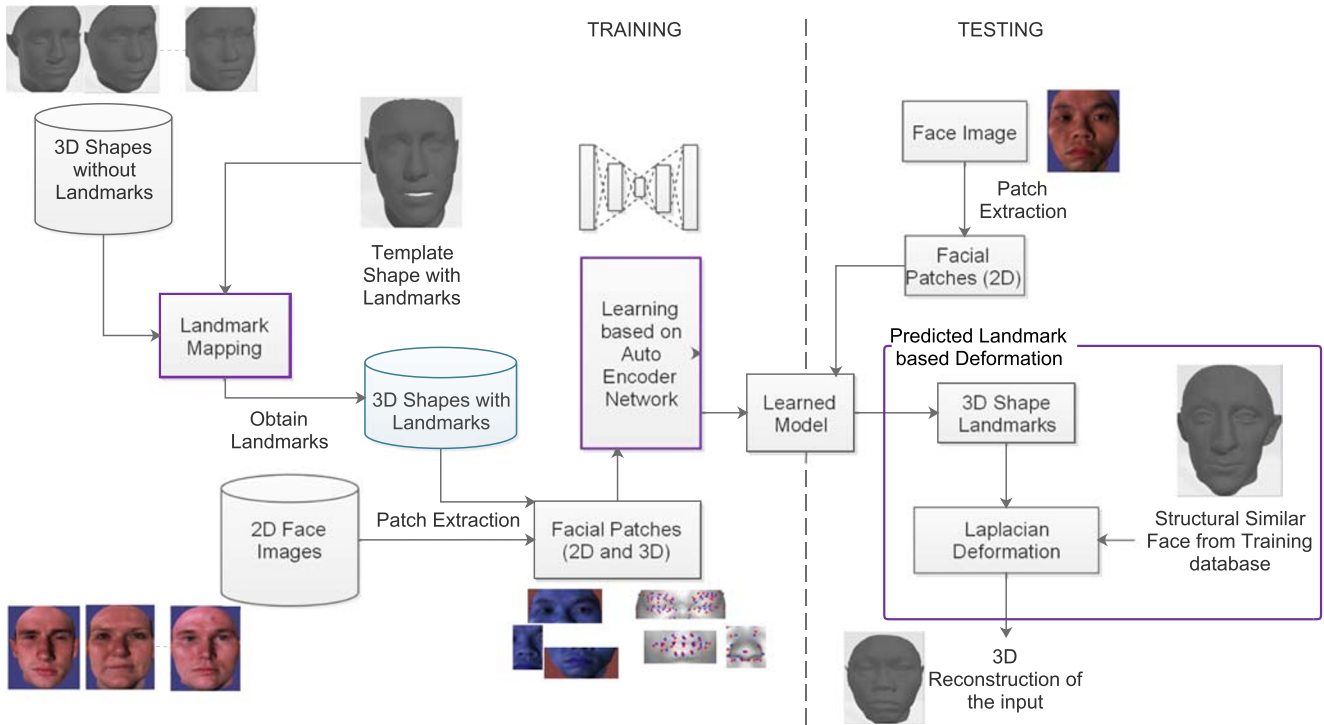
Fig. 1 The proposed SL2E-AFRE framework for 3D face reconstruction

divided into three partitions $P = \{P_e, P_n, P_m\}$, where $l_i \in \mathbb{R}^{3n}$; P_e, P_n and P_m are the sets encompassing eye patches, nose patches and mouth patches respectively. The edge map of the face image P_f is also used to estimate keypoints.

3.1 Enhanced keypoint set

An Enhanced keypoint set of 342 keypoints (Fig. 2) is built with intermediate serving keypoints for smooth deformation. It includes 68 base keypoints, 113 facial

contour vertices, 100 cheek vertices, 40 eye and eyebrow vertices, 20 nosebase vertices and 1 nosetip. In the z-axis, the largest valued vertex is taken as the nosetip vertex.

Since the target and reference shapes are aligned to each other, by exploiting the groundtruth landmark points from the reference shape, target shape's landmarks get inherited. The quantity of landmarks is expanded with the following landmark function:

$$\{v'\} = \{v\} \pm \Delta \tag{2}$$

Fig. 2 **a** 68 base keypoints, **b** Enhanced keypoint set
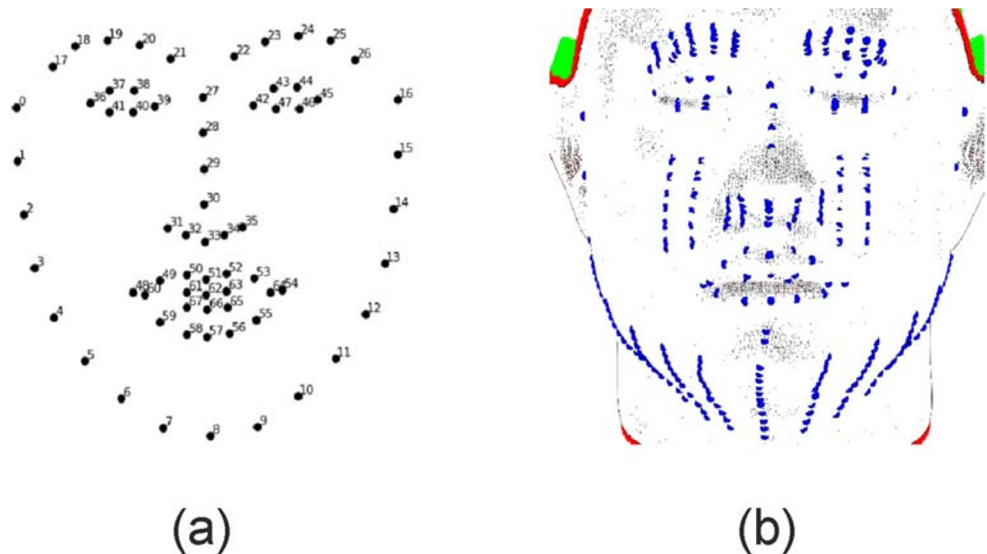
where $\{\nu\}$ is a set of preset landmark vertices; $\Delta = \{\delta_1, \delta_2, ..., \delta_n\}$. A small value δ_i is added and subtracted from $\{\nu\}$ to obtain a new set of landmarks $\{\nu'\}$. The set of supplementary eye landmarks is defined by,

$$\mathbb{L}_e = \{(\nu_e, \nu_e') \in \mathbb{V} \mid \nu_e' \text{ can be obtained from } \nu_e \text{ by applying (2)}\} \quad (3)$$

the set of supplementary nose landmarks is defined by,

$$\mathbb{L}_n = \{(\nu_n, \nu_n') \in \mathbb{V} \mid \nu_n' \text{ can be obtained from } \nu_n \text{ by applying (2)}\} \quad (4)$$

and the set of supplementary contour landmarks is defined by,

$$\mathbb{L}_c = \{(\nu_c, \nu_c') \in \mathbb{V} \mid \nu_c' \text{ can be obtained from } \nu_c \text{ by applying (2)}\} \quad (5)$$

where $\mathbb{V}$ is the vertices in the 3D shape model.

3.2 Landmark mapping across 3D databases

The proposed landmark mapping method is depicted in Fig. 3. The prerequisite to achieve 3D point cloud correspondence is that the shape of all the subjects(the topology of landmarks irrespective of the size of the subject) should be similar, so that all the subjects have equal number of vertices. In this paper, 3D facial landmarks are recognized with the use of 3dMDLab benchmark [7] as the reference. In some previous methods [28], locations of l_i is achieved with multivariate Gaussian distribution, i.e.,

$$p(y|x, \Theta) = N(1, \sigma^2) \quad (6)$$

$p(m_i, l_i | \Theta)$ is defined as,

$$p(m_i, l_i | \Theta) = p(m_i | l_i, \Theta) p(l_i | \Theta)$$
$$\propto exp\left(-\frac{||m_i - sURl_i - t||^2}{2\sigma_i^2} - \frac{||l_i - \mu_i||^2}{2\rho_i^2}\right) \quad (7)$$

where m_i represents the 2D frontal face feature points and Θ represents pose parameters $\Theta = \{s, R, t\}$, because a 2D face image can be formed by projecting the 3D shape geometry with Enhanced pose parameters such as rotation 'R', scaling

's', translation 't'. The 2D projection is carried out with $U_{2\times3}$=[1 0 0 ; 0 1 0] and the depth has not been included. Standard deviation σ and ρ are computed from the training dataset. And, $p(l_i|m_i, \Theta)$ is defined as,

$$p(l_i|m_i, \Theta) = \frac{p(m_i|l_i, \Theta) p(l_i|\Theta)}{p(m_i|\Theta)} \quad (8)$$

$p(m_i|\Theta)$ is represented by,

$$p(m_i|\Theta) = \int p(m_i, l_i|\Theta) dl_i \quad (9)$$

In this work, the landmarks are detected using a template model with landmark annotations. First, the mean shape of the USF database is rigidly aligned with the reference shape using the most widely used Iterative Closest Point(ICP) algorithm [5] for 3D shape registration. ICP aligns two moderately overlying meshes. This registration method establishes correspondence between the mean-USF and the reference triangulated mesh [1]. It first selects sample points from the target shape employing random sampling method, finds the closest points in the source mesh iteratively for 10 iterations employing knn search.

Then weights are provided to each correspondence based on (10).

$$W(m, b_i) = \frac{exp(-D(m, b_i))}{\sum_{i=1}^{N} exp(-D(m, b_i))} \quad (10)$$

where m is a vertex in the reference mesh, b_i is the ith closest vertex in the source mesh and $D(m, b_i)$ is the euclidean distance between m and b_i.

It uses sum of squared distance as the error metric to be minimized to derive final transformation.

$$vert\_aligned_{targ} = knn\_search(v_{targ}, v_{ref}) \quad (11)$$

where $vert\_aligned_{targ}$ is the aligned vertices of the target shape, which is the mean shape of USF database, v_{targ} is the vertices from the target shape, v_{ref} is the vertices from the reference shape. Then, the landmark points are recognized with the use of the reference 3D shape geometry landmarks.
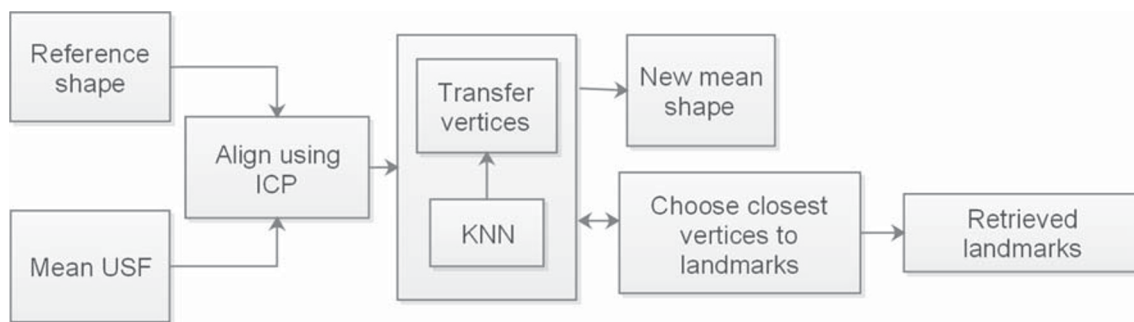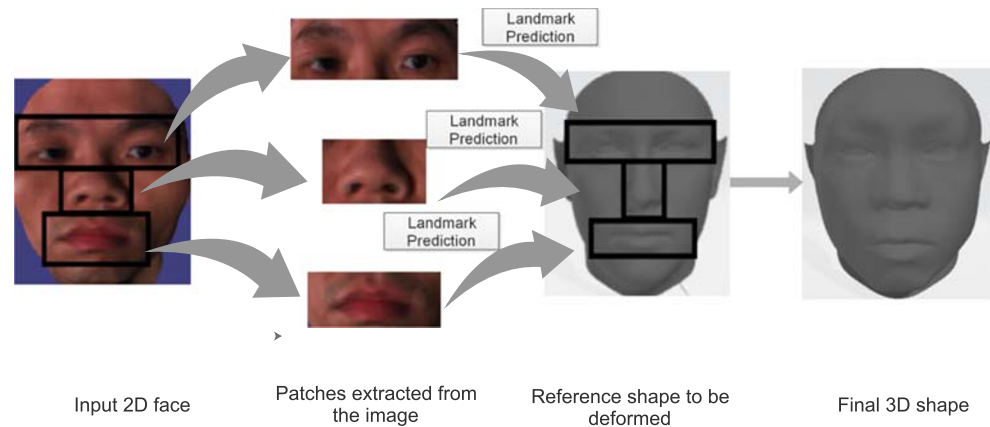

Fig. 3 Landmark mapping across 3D databases

Input 2D face   Patches extracted from   Reference shape to be   Final 3D shape
                            the image   deformed

3.3 Patch subspace learning & facial landmarks estimation with autoencoder

The autoencoder network for 3D landmarks estimation is shown in Fig. 7, in which, each layer is a fully connected layer. It comprises stacked dense layers. To assist the 3D face shape reconstruction, the input 2D face image is partitioned into three patches and the edge map of the face region is extracted.

3.3.1 Face partitioning

The proposed approach starts with partitioning the input faces into patches: eye, nose, mouth. Figure 4 shows how the patches are extracted and utilized to derive the landmarks for generating the final 3D geometry. Each facial patch with the predefined dimension is given as one dimensional input to the autoencoder network.

3.3.2 Edgemap for chin-cheek contour keypoints prediction

We can extract the chin-cheek region keypoints with the help of Sobel operator by detecting horizontal edges using the threshold value as 0.01. Edge detection operation depends on the quality of image. Therefore, preprocessing of image and post processing of edgemap should be performed accordingly. Here, the input image is converted into gray image followed by image dilation, which

makes the edges sharper. Then, edges are detected by applying sobel edge operator. Edge detection is followed by morphological close operation and area opening operation, which removes all the unwanted small edges consisting of less than a specific number of pixels. Unwanted eye and mouth region are removed through hole filling and deleting connected components with significant number of pixels. Finally, contour is extracted through morphological operations, namely dilation, extraction of largest blob and erosion. Only the lower half of the resultant image is analysed for predicting the required chin-cheek contour keypoints. Though we obtain similar result when other edge operators such as prewitt and canny are used, noise suppression characteristics of sobel is better than prewitt and sobel is computationally less expensive compared to canny [22]. The output after applying sobel method is shown in Fig. 5. The proposed method benefits from this edgemap by obtaining more accurate chin-cheek contour keypoints, resulting in an accurate facial shape.

3.3.3 Subspace learning and landmark estimation

After acquiring sufficient landmarks on the database faces using the proposed landmark mapping strategy and dividing each face into patches P_e, P_n, P_m and P_f, each patch set is given as input for the concerned autoencoders, since each category of patches is handled by different autoencoder networks separately. A simple autoencoder is shown in
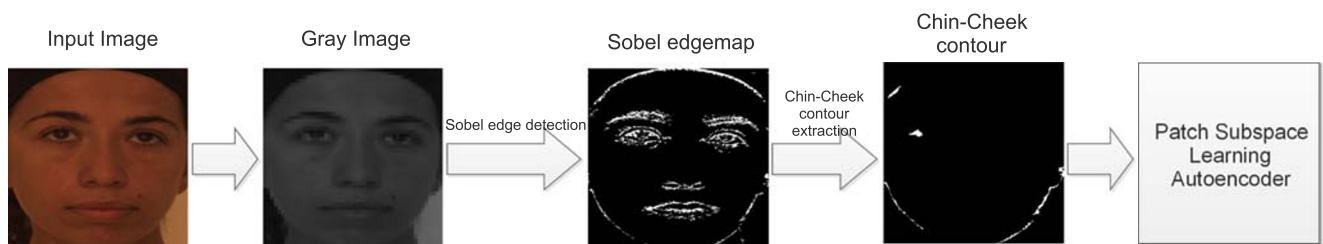
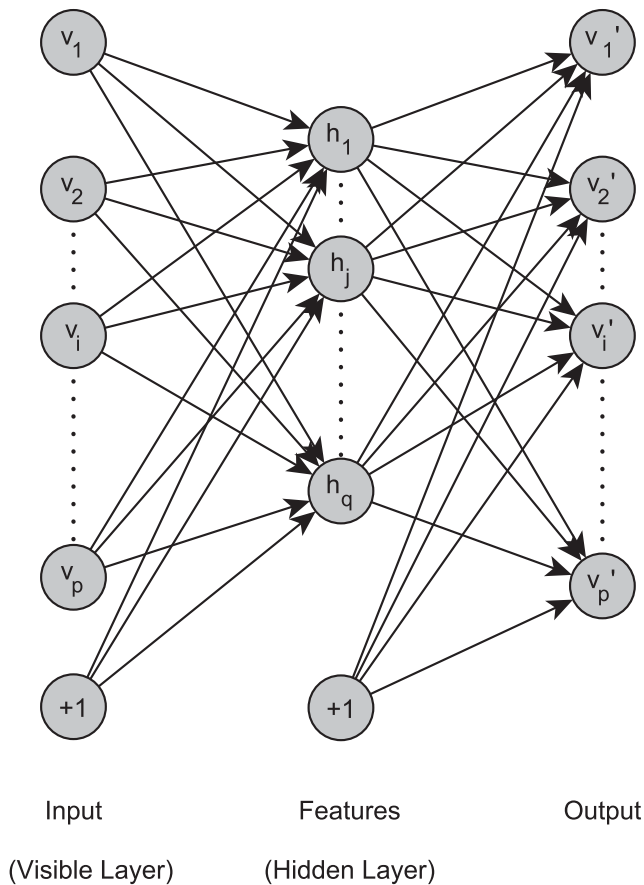**Fig. 5** Contour Extraction with Sobel

Input   **Features**   **Output**

(Visible Layer)   **(Hidden Layer)**

Fig. 6 Layers in a simple Autoencoder

Fig. 6. An autoencoder [3, 17] has two partitions: an encoder and a decoder. The patches are encoded using the 11-layered autoencoder depicted in Fig. 7. Here, the input is the pixel values of the pixels in each patch image. A

patch p_i is encoded using the scaled exponential linear unit activation function. The decoder reconstruct the input p_i from the encoded input. Generally, an autoencoder network is trained, such that the mean squared error between actual input(x_i) and the obtained output(x_i') gets minimized, since the expected output is nothing but the input itself.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}||x_i - x_i'||^2 \tag{12}$$

where N indicates the total number of objects. The autoencoder network parameters are obtained by minimizing the mean squared error.

In the proposed approach, the objective is minimizing image reconstruction error and landmark error simultaneously. The proposed energy function as specifed in (1) is,

$$\min_{W,b} \quad \frac{1}{N}\sum_{i=1}^{N}||p_i - p_i'||^2 + \alpha\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{p_i} - \hat{y}_{p_i})^2}$$
$$-\beta||enc(p_i) - c_i||_2 \tag{13}$$

where N is the size of the dataset, y_{P_i} and $\hat{y}_{P_i}$ are actual and obtained keypoints for the patch p_i; enc(p_i) and p_i' are encoded and decoded version of input face patch, c_i is the centroid to which p_i belongs. Since the landmark prediction relies on the clustering performance, which depends on the code layer output and the autoencoder does not provide the same encoded representation for the identical input, $cluster_{loss}$ is subtracted from the computed error. $cluster_{loss}$ is obtained by taking L_2 distance between the encoded input and the centroid it belongs. α and β are meta-parameters used to help estimating the model parameters and are set by trial and error.
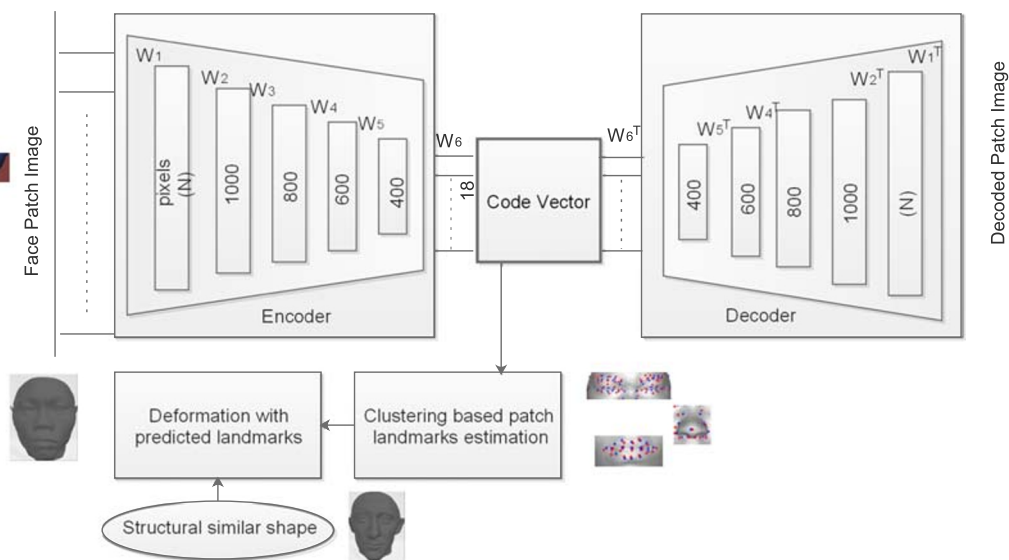

Fig. 7 The autoencoder network for 3D landmarks estimation

To compute the loss, in every iteration, clusters have been computed based on the 2D subspace dimension at the code layer. During the first iteration centroids are selected randomly. Each facial patch is assigned to a patch cluster centroid in every iteration.

$$c_i = \min_{c_k} \quad ||enc(p_i) - c_k||_2 \tag{14}$$

Thereby, the centroids get updated for the next iteration.

$$c_k{}^{t+1} = \frac{\sum_{\mathbf{p_i} \in C_k{}^t} enc(p_i)^t}{|C_k^t|} \tag{15}$$

where c_k is k^{th} centroid and C_k is the set of patches belonging to c_k. The algorithm for subspace learning and landmark estimation is shown in Algorithm 1.

Algorithm 1 2D Patch subspace learning and landmark estimation using autoencoder.

Input : 2D patches $P = \{P_e, P_n, P_m, P_f\}$, landmarks of the training 3D subjects L, meta-parameters $\{\alpha, \beta\}$, number of iterations I, number of clusters.

1 Training:
2 **for** *each patch set* **do**
3   Assign centroid randomly
4   iter ← 1
5   **while** *iter≤I* **do**
6   Find the code vector
7   Find the closest centroid of the patch
8   Retrieve the 3D landmarks of the subjects using Algorithm 2
9   Optimize the (13) with stochastic gradient descent method
10   Update centroids and centroid assignments for each subject's patch via (14)
11   iter ← iter+1
12   **end**
13 **end**
14 Testing:
15 **for** *each patch set* **do**
16   Find the code vector
17   Find the closest centroid of the patch
18   Retrieve the 3D landmarks of the subjects using Algorithm 2
19 **end**

To find the relevant samples among the cluster of samples, SSIM(Structural Similarity Index Measure) is computed between the input and each of the cluster samples. SSIM can be defined as,

$$SSIM(p, q) \quad = \quad [l(p, q)^\alpha . c(p, q)^\beta . s(p, q)^\gamma] \tag{16}$$

where, α, β, γ are weights given to each of the comparative measures luminance(l), contrast(c), structure(s).

$$l(p, q) = \frac{2\mu_p \mu_q + C_1}{\mu_p^2 + \mu_q^2 + C_1} \tag{17}$$

$$c(p, q) = \frac{2\sigma_p \sigma_q + C_2}{\sigma_p^2 + \sigma_q^2 + C_2} \tag{18}$$

$$s(p, q) = \frac{\sigma_{pq} + C_3}{\sigma_p \sigma_q + C_3} \tag{19}$$

With, $C3 = C2/2$ and setting α, β, γ to 1, the equation can be reduced to,

$$SSIM(p, q) = \frac{(2\mu_p \mu_q + C_1)(2\sigma_{pq} + C_2)}{(\mu_p^2 + \mu_q^2 + C_1)(\sigma_p^2 + \sigma_q^2 + C_2)} \tag{20}$$

Then, based on a threshold value ϵ, a subset is selected from the cluster, named as enhanced cluster M_c. For a patch p_i, M_c is computed as follows:

$$M_c = \{s = [s_1, ...s_m, ...s_n]; s \subset C_k, s_m \in C_k \wedge SSIM(p_i, s_m) < \epsilon\} \tag{21}$$

where C_k is the cluster comprising the samples belonging to the cluster centroid c_k. The cluster subset is used to estimate the facial patch landmarks.

$$\hat{y}_{p_i} = \frac{1}{|M_c|} \sum_{\mathbf{s_m} \in M_c} y[s_m] \tag{22}$$

where $|M_c|$ is the number of samples in M_c and $y[s_m]$ is the keypoints of s_m. It is explained in the Algorithm 2.

Algorithm 2 Landmark estimation.

Input : chosen cluster C_k, threshold ϵ
1 **for** *each element z_i in C_k* **do**
2   Compute SSIM using (20)
3   **while** *SSIM<ϵ* **do**
4   Insert the patch to the enhanced cluster $\{M_c\}$
5   **end**
6 **end**
7 Compute the landmarks with the $\{M_c\}$ using (7)

In each iteration, the mapping function and the landmark loss function need to be optimized (13) via back-propagation.

Upon completion of the learning process, the code layer dimension(5th layer output) is utilized as feature vector for the new test input face patch. Based on its closest centroid and (21), landmarks are predicted. The patch landmarks are combined together to obtain the complete face landmarks.

After the facial landmarks estimation, laplacian coordinate based surface deformation is carried out by deforming a reference model, which is structurally similar to the predicted keypoints.

3.4 Predicted landmark based deformation

Since the resultant 3D shape achieved through deforming the mean shape lags in the fine details, predicted landmark based deformation is proposed here, in which the depth of non-landmark vertices are assigned to the most similar 3D face according to the predicted keypoints. The reference 3D face geometry is chosen in accordance with the minimum gap between the predicted 3D landmark vertices and the reference shapes' 3D landmark vertices. Since the 2D coordinates are different, first the reference landmark points get interpolated at the predicted 2D coordinates to reproduce the depth value. Thereafter, the distance is measured with (23), using the difference between the interpolated depth value and the predicted depth value.

$$distance = \sqrt{\|depth_{inter} - depth_{predict}\|} \qquad (23)$$

where $depth_{inter}$ is the reproduced depth value after interpolation and $depth_{predict}$ is the predicted landmarks' depth coordinate values. Then, the selected reference 3D model is deformed with the predicted landmarks using laplacian deformation method with Iterative Closest Point(ICP) algorithm.

4 Experimental results

This section presents the information about the databases used, the results and the comparative analysis which shows the effectiveness of the system that has been proposed.

4.1 Datasets

To evaluate the proposed methodology, experiments are implemented with two datasets. The Bosphorus 3D Face Database [31] includes 105 subjects with different expressions. In this database, facial information are captured by structured-light based 3D digitizer and the resolution of the 2D pictures are high (1600×1200 pixels). The resolution in x, y and z coordinates are 0.3mm, 0.3mm and 0.4mm respectively and this database incorporates six universal expressions (Happy, Sadness, Surprise, Disgust, Angry, Fear). In this experiment, neutral expression has been considered. The USF Human-ID 3D Face database [6] consists of 100 laser scanned 3D faces, scanned under controlled viewing conditions and each subject in the database has 75972 vertices. The 2D face images were obtained by taking

Fig. 8 Examples of the 3D shapes obtained by deforming the mean USF using the landmarks predicted with the proposed landmark mapping strategy. The first column is the face images, the second column shows the 3D shapes from the database, the third column shows the deformed 3D shapes using the landmarks predicted, the fourth column shows the error map(*Blue-Green-Yellow-Red*)
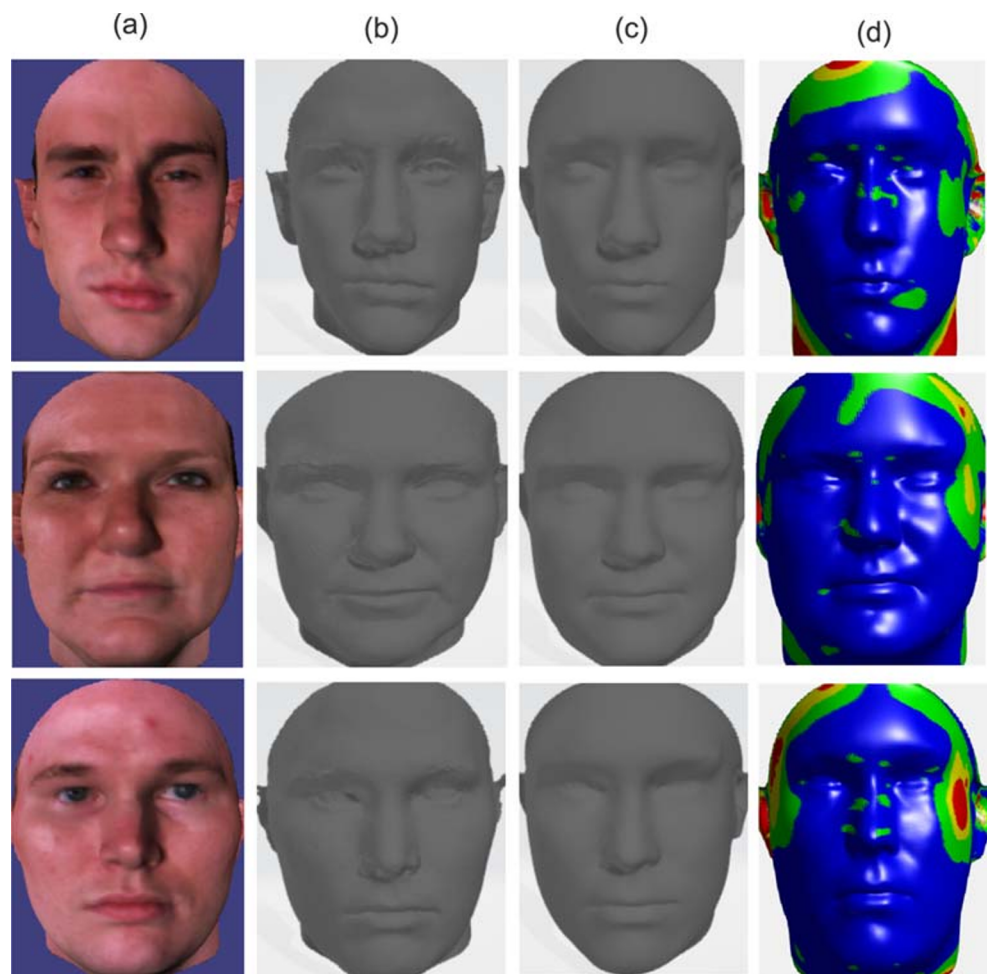
Table 1 Performance evaluation of ISK

Intermediate	USF Database		Bosphorus Database	
serving keypoints	Avg. reconstruction error	Avg. RMSE	Avg. reconstruction error	Avg. RMSE
Without ISK	4.17±0.51	0.187	4.13±0.81	0.092
With ISK	3.86±0.69	0.132	3.05±1.02	0.043

snapshots of the 3D face geometry. 3dMDLab benchmark [24] used in landmark mapping purpose provides around 68 keypoints locations. In addition, intermediate serving keypoints are found by adding and subtracting small numbers of consecutive vertices as in (2), from important features among the 68 landmarks. The proposed autoencoder based 3D face reconstruction method is implemented using python on a 64-bit Windows operating system over Intel(R) Core(TM) i5-8300H processor @ 2.30GHz with NVIDIA Geforce GTX 1060 GPU and 8.0 GB RAM.

4.2 Results

4.2.1 Landmark detection

Manual landmark localization is most probably error prone and tedious task. So, the 3D landmarks are found with the use of the mean shape of some publicly available faces from 3dMDLab benchmark [7] and the annotated landmarks. Initially, 68 primary and secondary keypoints are found with the proposed landmark mapping method, and then, in addition, 274 intermediate serving keypoints are generated from the obtained keypoints by adding and subtracting

small numbers of consecutive vertices. Figure 8 shows some examples of the 3D shapes obtained by deforming the mean USF using the landmarks predicted with the proposed landmark mapping strategy.

This 3D face reconstruction work represents a valuable alternative to the state-of-the-art conventional [23, 44, 46] and deep neural network learning based [2, 21, 24, 33] methods. In the former, shading information and sophisticated algorithms are used and in the later, 2D face subspace as well as the corresponding 3D shape geometry are learned with computationally expensive networks. However, the proposed work learns only 2D face subspace without such non-trivial networks. In addition, intermediate serving keypoints carry contour and cheek shape details.

### 4.2.2 3D face reconstruction comparison with and without ISK

The performance of the ISK on the proposed SL2E-AFRE method is evaluated. First, 3D reconstruction is carried out with the 68 base keypoints. With them, the facial shape could not be recovered more accurately. For that, a new enhanced keypoint set including base as well as intermediate serving keypoints is introduced and used for the proposed approach. The Table 1 summarizes overall performance of the introduced ISK. As the number of important keypoints increases, the system generates more accurate 3D geometry.

4.2.3 Optimal choice of number of clusters

The optimal choice of the number of clusters to be used influences the performance to some extent. Experiments

Fig. 9 Impact of choosing the optimal number of clusters
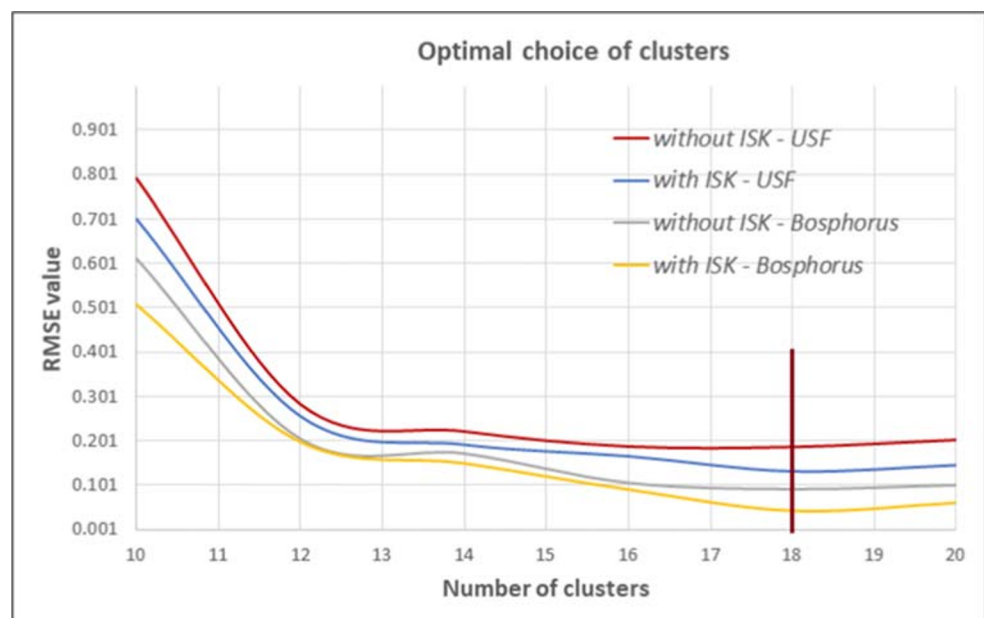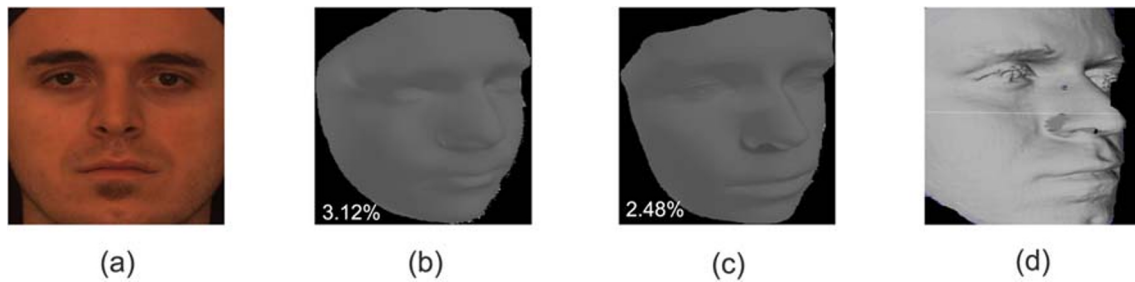
Fig. 10 **a** Input image, **b** Reconstruction with generic face, **c** Reconstruction with landmark based reference face, **d** Ground truth.

are carried out to determine the optimal number of clusters that both the datasets impart better performance. Figure 9 showed the number of clusters versus RMSE value. The proposed algorithm works efficiently at the value of 18 and so the number of clusters for the experiment is set to 18. Compared to the method without ISK, the method with ISK performed better on two datasets.

4.2.4 Deformation on generic face vs landmark based reference face

As shown in Fig. 10, when the generic face is used as the reference for deformation, it's semblance will be there in all the reconstructed models. Also, unique appearance details will be missed out. Hence, the landmark based reference model, which is structurally similar to the test subject is used for the deformation leading to a personalized unique 3D model. In the figure, the respective reconstruction errors are overlaid.

4.3 Comparison

4.3.1 Experiments on USF human-ID 3D face database

First, the proposed method is evaluated on USF Human-ID 3D Face database. Since, the less availability of large 3D face database is a big constraint in 3D face reconstruction task, training and testing is carried out in 90:10 ratio. The training set is further partitioned into train and validation sets. To deal with the over fitting problem, the samples for validation are randomly selected at each epoch. The meta-parameters has been set manually in (13) as $\alpha = 3.2 \times 10^{-3}$

Table 2 Face patches and their dimensions

Patch name	Dimension
Eye	50×23
Nose	20×30
Mouth	50×21
Face contour	50×50

and $\beta = 5 \times 10^{-3}$. The evaluation is carried out with the neutral expression as the baseline. The resolution of the input patches are given in the Table 2.

The number of clusters is set to 18 for both the dataset. To derive the enhanced cluster set for each patch, the threshold value $\epsilon = 0.63$ is used, as detailed in Algorithm 1. To evaluate the estimated landmarks, the reconstruction error [23] is employed to be the performance metric and is computed by,

$$\text{reconstruction error} = \frac{100.|(z_{rec} - z_{truth})|}{z_{truth}} \quad (24)$$

where z_{rec} is the z-coordinate value of the reconstructed shape keypoints and z_{truth} is the z-coordinate value of the ground truth keypoints.

The observable outcome of the proposed framework on some human faces from the USF Human-ID 3D Face database can be seen in Fig. 11. From the reconstruction error overlaid on the obtained 3D shapes, it is understood that the results are significant.

The proposed method is compared with [21, 23, 44, 46] and [2]. The method proposed by Kemelmacher et al. [23], in which one reference model is used for generating 3D models of all the test subjects resulting in reference model resemblance in the final model. The compared methods have been put forward with computationally very expensive deep networks [30], but the proposed approach is able to provide plausible result without such complex architectures.

We adopt the mean and standard deviation of the reconstruction error to compare the performance, which is detailed in Table 3, in which the best reconstruction is displayed in boldface. From the table, it is inferred that, SL2E-AFRE gives superior performance in obtaining the chin keypoints, which leads to better facial shape. There are still some failure cases due to the dataset constraint, since the proposed method relies fully on the keypoints.

4.3.2 Experiments on Bosphorus database

When the proposed approach is evaluated on the Bosphorus database, 90 out of 105 subjects are selected randomly and employed for training and validation purpose. And, the

Fig. 11 Sample Results. **a** The 2D image, **b** Groundtruth profile view, **c** Reconstructed profile view, **d** Groundtruth front view, **e** Reconstructed front view. The reconstruction error has been overlaid on each reconstructed face
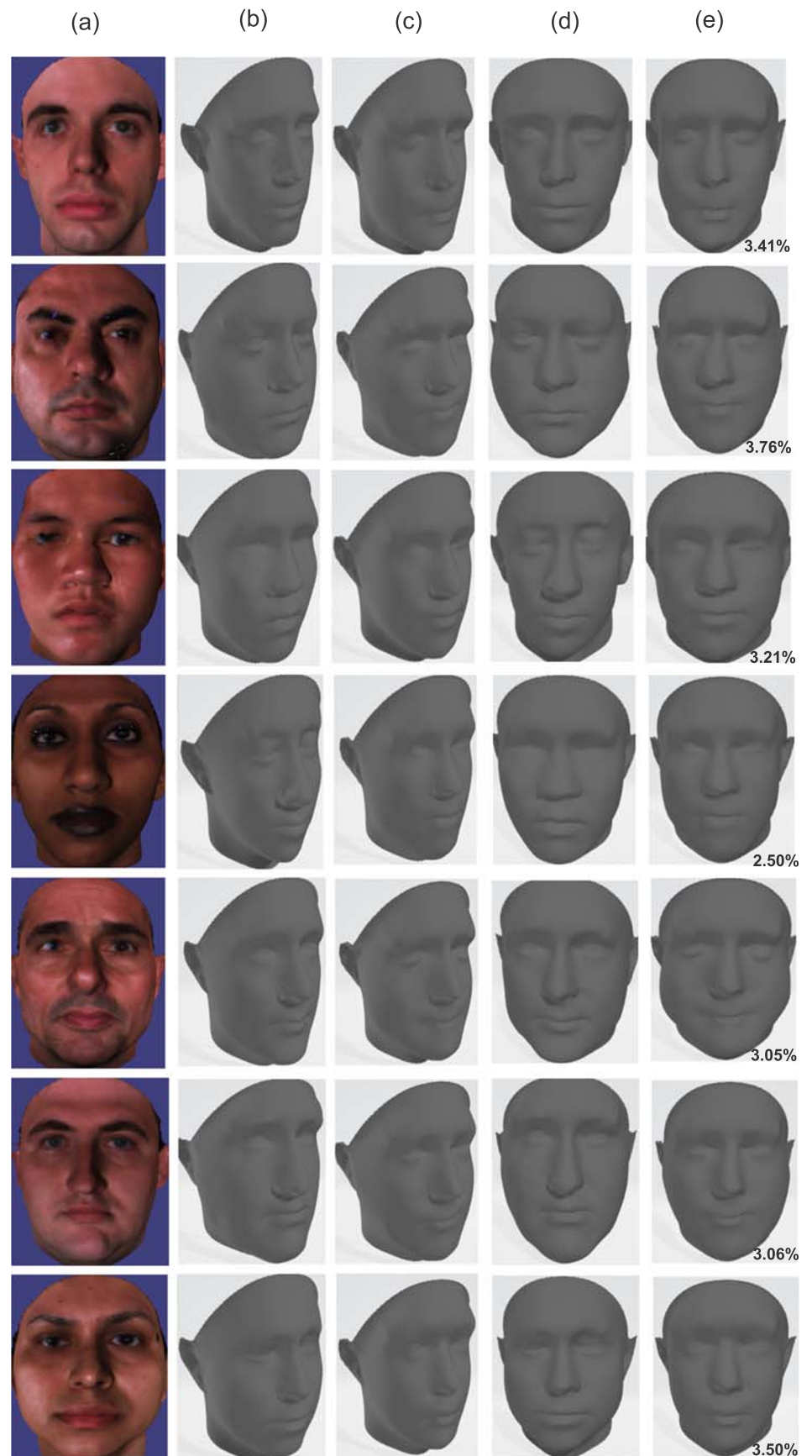
Table 3 Comparison of the mean and standard deviation of the reconstruction errors on the USF database

| Subjects	Face Patch	SFS	Zeng et al	Jiang et al	WGAN	Wu et al	SL2E-AFRE
Subject 1	Eye	6.80±2.50	5.91±1.52	5.20±1.02	4.98±0.02	**4.21±0.11**	4.25±0.21
	Nose	4.32±1.82	4.12±0.91	4.06±0.18	4.02±0.10	4.05±0.01	**3.11±1.10**
	Mouth	6.02±1.92	6.00±0.18	5.16±0.12	**4.18±1.21**	4.21±1.16	4.20±0.09
	Contour	3.19±1.09	2.89±1.06	3.18±1.01	2.82±0.10	3.01±0.10	**2.51±0.11**
Subject 2	Eye	6.85±1.89	5.82±0.68	4.89±0.02	4.88±0.01	4.85±1.02	**4.31±1.10**
	Nose	4.19±1.02	5.02±1.02	4.07±1.01	4.06±0.90	4.10±1.01	**3.12±1.28**
	Mouth	5.99±2.08	4.08±1.86	4.01±0.09	4.00±0.07	4.12±1.11	**3.91±0.30**
	Contour	4.12±1.78	4.98±0.82	4.11±0.16	4.17±1.12	4.11±0.16	**4.10±0.10**
Subject 3	Eye	7.21±0.89	6.01±1.03	6.08±1.12	4.91±0.12	5.01±1.07	**4.52±1.80**
	Nose	5.18±1.65	5.01±0.18	5.08±0.11	**3.00±0.16**	4.91±1.02	3.02±1.70
	Mouth	6.81±1.34	5.88±1.20	5.81±1.02	4.08±1.52	4.11±0.16	**4.02±1.50**
	Contour	5.81±2.18	4.09±1.18	4.11±0.08	2.91±0.12	2.17±0.01	**2.12±0.08**
Subject 4	Eye	5.91±1.56	4.19±0.88	3.91±1.50	3.71±0.11	**3.19±1.01**	3.20±1.50
	Nose	4.22±2.18	4.09±1.08	3.08±1.17	3.01±0.08	**2.17±1.91**	2.90±1.21
	Mouth	6.18±1.00	5.18±0.76	4.99±2.01	**4.01±0.12**	4.09±0.05	4.12±1.30
	Contour	3.71±1.81	3.70±0.18	4.01±0.16	2.50±0.87	2.90±0.14	**2.20±0.10**
Subject 5	Eye	6.12±1.18	5.12±2.12	5.91±0.88	4.33±0.06	**4.01±1.12**	4.32±1.72
	Nose	5.12±2.71	4.97±0.65	4.88±0.01	3.28±0.18	4.01±0.01	**3.21±1.81**
	Mouth	4.72±2.72	4.03±1.09	3.99±1.76	3.61±0.61	4.00±0.02	**3.52±1.52**
	Contour	4.18±2.01	5.76±2.12	4.19±0.42	4.30±0.03	5.01±0.12	**4.28±0.72**
Subject 6	Eye	5.42±1.08	5.01±0.19	4.31±0.12	**3.01±1.21**	3.22±1.09	3.31±1.05
	Nose	5.19±0.19	4.91±1.71	4.11±1.11	4.21±0.51	**4.09±0.66**	4.25±0.15
	Mouth	6.12±1.01	5.11±1.87	4.32±1.34	**3.88±1.08**	4.11±1.12	4.15±1.10
	Contour	5.19±0.09	5.09±1.16	4.28±1.12	3.17±0.89	**2.98±0.99**	3.31±0.31
Subject 7	Eye	6.81±1.07	5.91±0.11	4.33±0.08	4.20±0.01	4.16±0.15	**3.11±1.20**
	Nose	5.74±2.04	5.01±0.21	**4.02±0.11**	4.52±0.71	4.55±0.70	4.20±0.11
	Mouth	5.17±1.08	5.07±0.17	4.65±0.06	**3.01±0.01**	3.89±0.11	4.01±0.08
	Contour	6.07±0.28	4.91±0.21	4.59±0.71	3.92±0.17	3.99±0.12	**3.91±0.11**
Subject 8	Eye	5.39±1.12	5.21±0.11	4.28±0.12	4.02±1.01	**3.91±1.11**	4.17±0.11
	Nose	5.26±0.17	4.19±1.00	4.11±0.06	**3.56±0.11**	3.67±1.12	3.81±0.90
	Mouth	6.09±2.11	5.11±0.91	4.91±1.04	**3.23±1.20**	4.01±0.07	4.16±0.17
	Contour	5.11±0.12	4.00±0.06	3.65±0.61	3.56±0.81	**3.09±0.11**	3.69±0.77
Subject 9	Eye	5.91±0.07	4.98±0.81	3.91±0.05	3.61±0.23	3.01±0.11	**2.66±1.00**
	Nose	5.22±0.12	4.19±0.77	3.97±0.11	3.89±0.06	**3.51±0.52**	3.71±0.11
	Mouth	5.08±0.21	4.92±1.07	3.67±0.20	3.08±0.17	3.11±0.71	**2.91±0.11**
	Contour	6.11±0.01	5.11±1.10	4.86±0.45	3.96±1.30	3.91±1.20	**3.01±0.01**
Subject 10	Eye	5.39±0.66	4.38±1.02	4.41±0.06	3.01±0.08	3.31±0.12	**2.76±0.16**
	Nose	4.98±0.08	4.67±1.01	4.03±1.00	3.90±1.01	3.23±0.81	**3.01±0.78**
	Mouth	4.91±1.12	4.11±1.26	3.96±0.21	3.76±0.81	3.80±0.17	**3.72±0.08**
	Contour	5.26±0.16	4.99±1.02	4.71±0.62	4.01±0.02	4.02±0.12	**3.91±0.19**
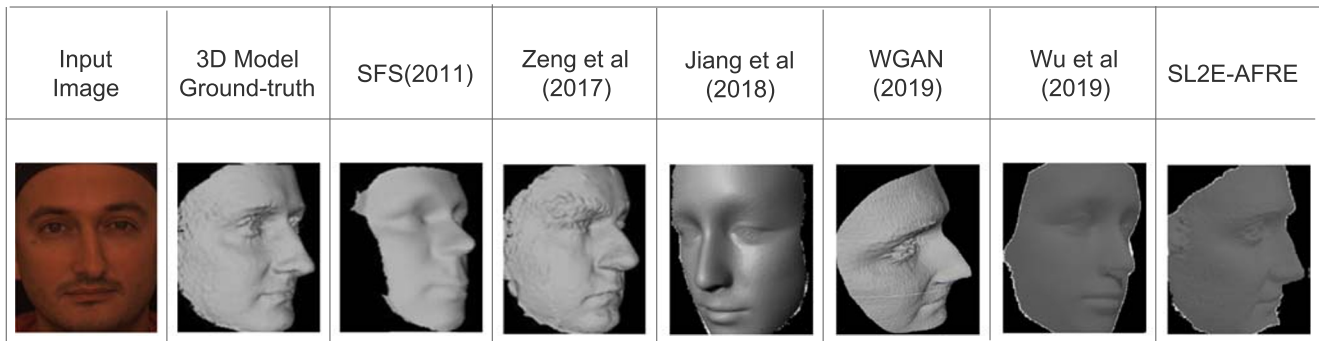
Fig. 12 Sample result of various state-of-the-art methods

subjects for the validation set are selected randomly in each epoch.

The predicted keypoints are employed to deform the selected 3D model from the training set of subjects chosen based on the landmark prediction. The proposed method is compared with [21, 23, 44, 46] and [2]. Kemelmacher et al. [23] used a single generic model for generating all 3D face models. However, the proposed approach uses different reference models for different facial patches, ensuring unique reconstruction and the RMSE for Kemelmacher et al. [23] is larger than that for the proposed approach. Zeng et al. [46] used both front and profile views of a face image, but the proposed method utilizes only frontal face. The methods proposed in [21] and [2] are deep neural network based methods. Arslan et al. [2] uses Generative Adversarial Networks(GAN), which incorporates both generator and discriminator network. However the result of the proposed approach is comparable to those computationally expensive methods. Figure 12 presents a comparison on a sample

input. From the Fig. 13, it can be seen that, the proposed method gives significant result on most of the subjects.

5 Conclusion and futurework

In this paper, a method for 3D face reconstruction with a single frontal face image is presented. Initially the keypoint localization in the databases is carried out using the landmark mapping approach. The proposed 3D reconstruction method employs different reference models for modeling different facial patches. It utilizes an autoencoder network for learning the patch subspace and selecting the reference models for landmark estimation. Finally, a reference shape is chosen from the database, which is considered as structurally similar to the test subject according to the predicted landmarks. Then, laplace deformation is carried out on the selected model with the predicted keypoints.
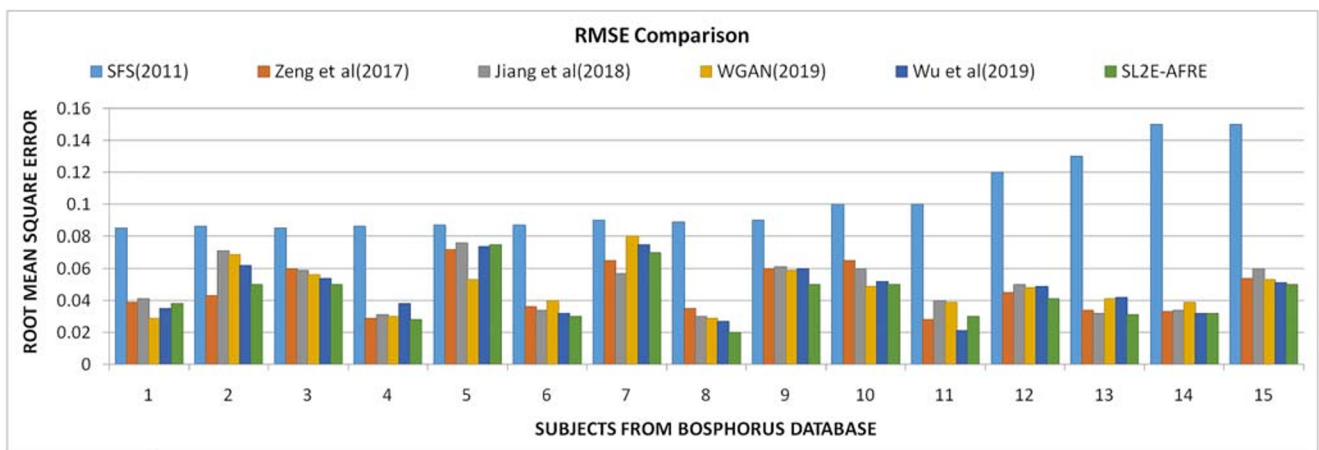

Fig. 13 Graphical representation of RMSE values computed on 15 subjects from Bosphorus database

Unlike existing approaches, the proposed method adopts a single face image as input and does not rely on a single reference model. The evaluation result has demonstrated that obtaining plausible 3D face geometry is possible without computationally expensive approaches. However the findings might not be generalized to view-invariant face images. Since the proposed method is kind of example based method, the performance depends on the faces in the training dataset.

Further studies, which take uncalibrated input images into account will need to be undertaken in future. Also, it is believed that this framework can be extended for generating 3D comic characters.

Compliance with Ethical Standards

Conflict of interest All the authors declare that they have no conflict of interest.

References

1. Amberg B, Romdhani S, Vetter T (2007) Optimal step nonrigid icp algorithms for surface registration. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
2. Arslan AT, Seke E (2019) Face depth estimation with conditional generative adversarial networks. IEEE Access 7:23,222–23,231
3. Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning, pp 37–49
4. Baumberger C, Reyes M, Constantinescu M, Olariu R, de Aguiar E, Santos TO (2014) 3d face reconstruction from video using 3d morphable model and silhouette. In: 2014 27th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, pp 1–8
5. Besl PJ, McKay ND (1992) Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, international society for optics and photonics, vol 1611, pp 586–606
6. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp 187–194
7. Booth J, Roussos A, Ververas E, Antonakos E, Ploumpis S, Panagakis Y, Zafeiriou S (2018) 3d reconstruction of "in-the-wild" faces in images and videos. IEEE Trans Patt Anal Mach Intell 40(11):2638–2652
8. Castelan M, Hancock ER (2004) Acquiring height maps of faces from a single image. In: Proceedings. 2nd international symposium on 3D data processing, visualization and transmission, 2004. 3DPVT 2004. IEEE, pp 183–190
9. Chang T, Li H, Wen G, Hu Y, Ma J (2019) Facial expression recognition sensing the complexity of testing samples. Appl Intell 49(12):4319–4334
10. Ding B, Wang Y, Yao J, Lu P (2006) A fast individual face modeling and facial animation system. In: International conference on technologies for E-learning and digital entertainment. Springer, pp 980–988
11. Ding L, Ding X, Fang C (2014) 3d face sparse reconstruction based on local linear fitting. Vis Comput 30(2):189–200
12. Dou P, Shah SK, 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
13. Feng Y, Wu F, Shao X, Wang Y, Zhou X (2018) Joint 3d face reconstruction and dense alignment with position map regression network. In: Proceedings of the European conference on computer vision (ECCV), pp 534–551
14. Gowsikhaa D, Abirami S, Baskaran R (2014) Automated human behavior analysis from surveillance videos: a survey. Artif Intell Rev 42(4):747–765
15. Han L, Xiao Q, Wang S (2016) 3d face reconstruction from a single frontal face image by robust cascaded regression. In: 2016 international symposium on computer, consumer and control (IS3C). IEEE, pp 841–845
16. Hartigan JA, Wong MA (1979) Algorithm as 136: a k-means clustering algorithm. J Royal Stat Soc Series C (Applied Statistics) 28(1):100–108
17. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
18. Horn BK (1975) Obtaining shape from shading information. Psychol Comput Vis: 115–155
19. Jackson AS, Bulat A, Argyriou V, Tzimiropoulos G (2017) Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. In: Proceedings of the IEEE international conference on computer vision, pp 1031–1039
20. Jiang D, Hu Y, Yan S, Zhang L, Zhang H, Gao W (2005) Efficient 3d reconstruction for face recognition. Pattern Recogn 38(6):787–798
21. Jiang L, Zhang J, Deng B, Li H, Liu L (2018) 3d face reconstruction with geometry details from a single image. IEEE Trans Image Process 27(10):4756–4770
22. Joshi M, Vyas A (2020) Comparison of canny edge detector with sobel and prewitt edge detector using different image formats. Int J Eng Res Technol (1):133–137
23. Kemelmacher-Shlizerman I, Basri R (2011) 3d face reconstruction from a single image using a single reference face shape. IEEE Trans Pattern Anal Mach Intell 33(2):394–405
24. Liang H, Liang R, Song M, He X (2016) Coupled dictionary learning for the detail-enhanced synthesis of 3-d facial expressions. IEEE Trans Cybern 46(4):890–901
25. Luo C, Zhang J, Yu J, Chen CW, Wang S (2019) Real-time head pose estimation and face modeling from a depth image. IEEE Trans Multimedia
26. Karthika Devi MS, Shahin Fathima RB (2019) Cbcs - comic book cover synopsis: Generating synopsis of a comic book with unsupervised abstarctive dialogue. In: International conference on 9th world engineering education forum 2019
27. Karthika Devi RB, Shahin Fathima MS (2019) Sync- short yet novel concise natural language description: Generatimng a short story sequence of an album images using multi modal network. In: International conference on ICT for sustainable development
28. Park SW, Heo J, Savvides M (2008) 3d face econstruction from a

single 2d face image. In: 2008 IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, pp 1–8

29. Patel NM, Zaveri M (2012) 3d model reconstruction and animation from single view face image. In: 2012 international conference on audio, language and image processing (ICALIP). IEEE, pp 674–682

30. Richardson E, Sela M, Or-El R, Kimmel R (2017) Learning detailed face reconstruction from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1259–1268

31. Savran A, Alyüz N, Dibeklioğlu H, Çeliktutan O, Gökberk B, Sankur B, Akarun L (2008) Bosphorus database for 3d face analysis. In: European workshop on biometrics and identity management. Springer, pp 47–56

32. Sivarathinabala M, Abirami S, Baskaran R (2015) View invariant human action recognition using improved motion descriptor. In: Computational intelligence in data mining, vol 3. Springer, pp 545–554

33. Song M, Tao D, Huang X, Chen C, Bu J (2012) Three-dimensional face reconstruction from a single image by a coupled rbf network. IEEE Trans Image Process 21(5):2887–2897

34. Sun Y, Jian M, Dong J (2016) Human face reconstruction from a single input image based on a coupled statistical model. In: Bio-inspired computing-theories and applications. Springer, pp 373–378

35. Tozza S, Falcone M (2016) Analysis and approximation of some shape-from-shading models for non-lambertian surfaces. J Math Imaging Vis 55(2):153–178

36. Tran AT, Hassner T, Masi I, Paz E, Nirkin Y, Medioni GG (2018) Extreme 3d face reconstruction: Seeing through occlusions. In: CVPR, pp 3935–3944

37. Tran L, Liu X (2019) On learning 3d face morphable model from in-the-wild images. IEEE Trans Pattern Anal Mach Intell

38. Wei W, Xu Q, Wang L, Hei X, Shen P, Shi W, Shan L (2014) Gi/geom/1 queue based on communication model for mesh networks. Int J Commun Syst 27(11):3013–3029

39. Wei W, Fan X, Song H, Fan X, Yang J (2016) Imperfect information dynamic stackelberg game based resource allocation using hidden markov for cloud computing. IEEE Trans Serv Comput 11(1):78–89

40. Wei W, Song H, Li W, Shen P, Vasilakos A (2017) Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network. Inform Sci 408:100–114

41. Wei W, Su J, Song H, Wang H, Fan X (2018) Cdma-based anti-collision algorithm for epc global c1 gen2 systems. Telecommun Syst 67(1):63–71

42. Wei W, Xia X, Wozniak M, Fan X, Damaševičius R, Li Y (2019) Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels. Comput Netw 161:210–219

43. Wei W, Zhou B, Połap D, Woźniak M (2019) A regional adaptive variational pde model for computed tomography image reconstruction. Pattern Recogn 92:64–81

44. Wu F, Li S, Zhao T, Ngan KN, Sheng L (2019) Cascaded regression using landmark displacement for 3d face reconstruction. Pattern Recogn Lett 125:766–772

45. Wu Y, Ji Q (2019) Facial landmark detection: a literature survey. Int J Comput Vis 127(2):115–142

46. Zeng D, Zhao Q, Long S, Li J (2017) Examplar coherent 3d face reconstruction from forensic mugshot database. Image Vis Comput 58:193–203

47. Zhang J, Zhuang YT (2007) Sample based 3d face reconstruction from a single frontal image by adaptive locally linear embedding. J Zhejiang University-SCIENCE A 8(4):550–558

48. Zhang J, Li K, Liang Y, Li N (2017) Learning 3d faces from 2d images via stacked contractive autoencoder. Neurocomputing 257:67–78

49. Zhou X, Leonardos S, Hu X, Daniilidis K (2015) 3d shape reconstruction from 2d landmarks: A convex formulation. In: Proceedings of IEEE conference on computer vision and pattern recognition. Citeseer, pp 4447–4455

P. R. Suganya Devi is a Research Scholar in the Department of Computer Science and Engineering, College of Engineering, Guindy, Anna University, Chennai, India. She has completed her graduation B.E. degree in Computer Science and Engineering in St.Xavier's Catholic college of Engineering, Anna University, Chennai. She has obtained her post graduation M.E. in Computer Science and Engineering from College of Engineering, Guindy, Anna University, Chennai. Her research interests include Image Processing, Face Recognition and 3D Reconstruction.

R. Baskaran is Professor of Computer Science and Engineering at College of Engineering, Guindy Campus, Anna University. He has received a prestigious Young Faculty Research Fellow (YFRF) in the year 2019 under the Visvesvaraya PhD Scheme under Ministry of Information Technology (Meity). He has more than 100 scholarly publications, including several volumes, spanning Databases, Data Mining, Software Engineering, Video Analytics, Artificial Intelligence, Machine Learning, Multimedia, Robotics, Web Services and related domains. He has supervised and currently supervising many PhD & MS scholars in the area of Computer Science and Engineering.