10-1-2020

3D Face Reconstruction from Single 2D Image Using Distinctive Features

H. M. Rehan Afzal
University of Newcastle

Suhuai Luo
University of Newcastle

M. Kamran Afzal
Xiamen University

Gopal Chaudhary
Bharati Vidyapeeth's College of Engineering

Manju Khari
Ambedkar Institute of Advanced Communication Technologies and Research

See next page for additional authors

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

3D Face Reconstruction from Single 2D Image Using Distinctive Features

H. M. Rehan Afzal[1], Suhuai Luo[2], M. Kamran Afzal[3], Gopal Chaudhary[4], Manju Khari[5], Sathish Kumar[6]

[1]School of Electrical Engg and Computing, University of Newcastle, Callaghan, NSW 2308, Australia. (email: hafizmuhammadr.afzal@newcastle.edu.au)
[2]School of Electrical Engg and Computing, University of Newcastle, Callaghan, NSW 2308, Australia. (email: suhuai.luo@newcastle.edu.au)
[33]School of Computer Science and Technology, Xiamen University, Xiamen, Fujian, 361005, China (email: kafzal@xmu.edu.cn)
[4]Bharati Vidyapeeth's College of Engineering, New Delhi, India (email: gopal.bvcoe@bharatividyapeeth.edu)
[5] Sr. IEEE Member, Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India. (e-mail: manjukhari@yahoo.co.in).
[6] Sr. IEEE Member, Dept. of Electrical Engg and Computer Science, Cleveland State University, Cleveland, OH USA (e-mail: s.kumar13@csuohio.edu).

Corresponding author Sathish Kumar (e-mail: s.kumar13@csuohio.edu).

ABSTRACT 3D face reconstruction is considered to be a useful computer vision tool, though it is difficult to build. This paper proposes a 3D face reconstruction method, which is easy to implement and computationally efficient. It takes a single 2D image as input, and gives 3D reconstructed images as output. Our method primarily consists of three main steps: feature extraction, depth calculation, and creation of a 3D image from the processed image using a Basel face model (BFM). First, the features of a single 2D image are extracted using a two-step process. Before distinctive-features extraction, a face must be detected to confirm whether one is present in the input image or not. For this purpose, facial features like eyes, nose, and mouth are extracted. Then, distinctive features are mined by using scale-invariant feature transform (SIFT), which will be used for 3D face reconstruction at a later stage. Second step comprises of depth calculation, to assign the image a third dimension. Multivariate Gaussian distribution helps to find the third dimension, which is further tuned using shading cues that are obtained by the shape from shading (SFS) technique. Thirdly, the data obtained from the above two steps will be used to create a 3D image using BFM. The proposed method does not rely on multiple images, lightening the computation burden. Experiments were carried out on different 2D images to validate the proposed method and compared its performance to those of the latest approaches. Experiment results demonstrate that the proposed method is time efficient and robust in nature, and it outperformed all of the tested methods in terms of detail recovery and accuracy.

INDEX TERMS 3D face reconstruction, feature extraction, facial modeling, gaussian distribution.

I. INTRODUCTION

3D face reconstruction has significant importance in the field of computer vision and image processing. During the past few years, it has gained major attention due to the limited applications of 2D images. The major contribution of 3D images in numerous fields - like medical, robotics, and security - has drawn the attention of image processing researchers. 3D images give us opportunities to perform beyond the limitations that are associated with 2D images. 3D reconstruction is being used for various applications, but due to its accuracy and efficiency for matching of faces in 3D domain, face recognition is considered mostly challenging.

There are many existing methods of 3D facial reconstruction, of which a few typical ones are briefly mentioned here. Jiang et al. [1] developed an algorithm from unconstrained 2D images to reconstruct 3D faces using a strategy named smooth coarse to fine optimization. They generated 3D faces using smooth coarse by aligning the projection of 2D landmarks with 3D landmarks. Xiaoguang et al. [2] proposed a method - a 2D assisted, self-supervised learning method (2DASL) - that can be used to reconstruct 3D images effectively "in the wild" of 2D facial images. Ding et al. [3] presented a detailed survey on face recognition algorithms that consist of 2D images. They

concluded that 2D face recognition is sensitive to pose and illumination. Due to this drawback, researchers are investigating innovative solutions to address it. Another survey was presented by Bowyer et al. [4] highlighting the importance of 3D face recognition, which concludes that 2D face recognition limitations can be overcome by 3D face recognition. These surveys depict the significance of 3D face reconstruction. 3D face reconstruction has received attention over the past three decades, since Hu et al. [5] reconstructed a 3D surface using the shape from shading technique. The method is based on the principle that with the changing of surface orientation, shading also changes. The change in orientation of surface helps to recover the surface based on intensity of image pixel. Before solving the problem of shape from shading (SFS), image formation must be known.

Many researchers have shared their ideas on the subject, but the pioneers of SFS are Horn and Pentland. SFS deals with the face as a Lambertian surface (LS). More precisely, the Lambertian surface's apparent brightness is same for each observer, irrespective of the observer's angle of view. For more understanding, readers can consult [6]. Later, using this concept, Queau et al. [7] also implemented the shape from shading technique to reconstruct a 3D image. These techniques have drawbacks due to several limitations. The first drawback is that these techniques can be applied only on Lambertian surfaces. Secondly, depth finding for the images that have specular surfaces, is not possible due to different reflectance behavior of that surface. Specular surfaces are mirror-like surfaces. So, depth finding is a very difficult process, and results are even worse for synthetic data. There is a need of such method which can work on every type of surface.

After the SFS era, researchers completely changed the method by introducing statistical learning-based techniques [6-8]. Statistical learning-based methods follow the rules to learn data from information. Many researchers proposed different methods. Zhang et al. [8] were one of them; they proposed a method on temporal-based correspondence, which had good results at high resolution but demanded a complex setup with controlled illumination - which is computationally expensive. But again, these methods have drawbacks as well as: they need multiple input images to reconstruct each 3D image. These methods also need intensive human interference for tracking the facial features. On the contrary, some methods [9-15] are no longer required to mark these features manually but instead use multi-view stereo methodologies, which require controlled lighting and calibration of cameras. Jo et al. [16] proposed a method for 3D face reconstruction that is an example of these algorithms. However, the method also needs manual annotations for the features of facial points. For a novice user, it is very difficult to mark these features, which shows that these techniques are laborious. One more method of 3D face reconstruction using multiple images was proposed by Kemelmacher Shlizerman and Seitz [17]. The method is based on a photometric stereo that produces high quality models of faces from multiple photos, but the locally consistent shape's recovery is attained by the collections of separate subsets of images. Neural networks also follow the same path of the learning-based technique. It modernized the applications of computer vision. These methods have good performance for regression problems. They use huge collections of images which are computationally expensive. Most of time, unless a large dataset is available, neural network methods don't provide promising results. An example of face reconstruction using neural networks is three-dimensional movie maker (3DMM) [18], whose limitation is that it uses nearly frontal images only.

To deal with pose estimation using frontal images and pose correction, researchers use probability techniques [19-22] in these learning-based methods. A novel method by Zhou et al. [23] is an example of these efforts. They applied an expectation maximization (EM) algorithm with a 3D face-deformable model for pose estimation. This method is appropriate for both types of poses: profile and frontal. The work of Vlasic et al. [24] is also related to the family of EM algorithm. They proposed a model that is multi linear in nature. A three-dimensional face dataset was used that contained different identities and expressions. They divided it into different modes consisting of expression, identity and viseme. 15 identities with ten expressions were estimated. For the missing data, they used an expectation maximization approach. Later, Chen et al. [25] proposed such an example by exploiting the face symmetry. The main advantage of this algorithm is that due to symmetry, it uses only one single image in order to reconstruct the 3D surface instead of multiple images. These new methods consisting of morphable models are a good addition to the chain of methods to reconstruct a 3D face. This idea was first used in telecommunications for the compression of images. Then it was introduced by the computer vision researchers. The main concept is that researchers are able to synthesize the 2D image with a specific model to reconstruct the a3D image. Volker Blanz et al. [26] proposed a method by implementing this idea to generate a 3D model of a face. They introduced texture and shape vectors, and then linearly combined the model of the 3D face by generating a 3D shape model prototype. A single image was used for reconstruction of the 3D image, which is computationally efficient. Our proposed method also follows the morphable model technique, and overcomes the problems discussed behind and described in contributions.

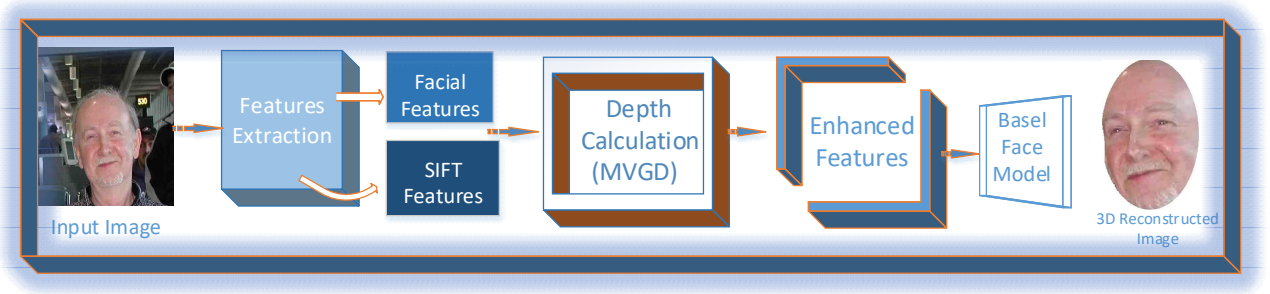


FIGURE 1. General structure of the proposed method

II. CONTRIBUTIONS

One of the main contributions of our proposed method is that it is computationally efficient because it takes only a single image to reconstruct a 3D image efficiently. Due to less expensive computations and processing speed, it can be implemented in real system. The method's effectiveness can also be seen in the results presented in section 4 by comparing its root mean square error (RMS) with those of the existing state-of-the-arts methods.

The second contribution of the proposed method is that, it not only takes facial features, but also extracts distinctive SIFT features, which results in an efficient 3D reconstruction. Distinctive features help to align images easily with a Basel face model. Thirdly, depth is calculated by Gaussian multivariate distribution, and then enhanced details are extracted by shape using shading technique, making the results optimal compared to original images and at the end achieving an accurate 3D face alignment. We need to point out that some algorithms discussed above have limited rotation capability and cannot rotate face up to 90 degrees. On the contrary, the proposed method can rotate the face at any angle, even at 90 degrees or side view, which can be seen in the evaluation section. Another important contribution of our method is that the process to generate a 3D face is fully automatic. It doesn't need any user-computer interaction as most existing methods do. Even a novice user with less computer knowledge can reconstruct a 3D image. These contributions make our algorithm efficient and effective compared with previous methods.

The block diagram of the proposed method can be seen in Fig. 1. The rest of this paper is summarized as below. Section 3 explains the proposed method which is further comprised of features extraction, depth calculation, and creation of 3D images. Results are evaluated in section 4 by comparing existing methods. Finally, the conclusion is explained in section 5.

III. THE PROPOSED METHOD

This section summarizes the proposed method. The method takes a single 2D image as input, finds the face in the image, and then performs different operations (discussed below) and gives a 3D reconstructed face as shown in Fig. 1. Our method consists of three steps: features extraction, depth calculation, and creation of 3D image. Feature extraction is sub-categorized with facial feature calculation and SIFT feature calculation. Depth calculation is performed with Multivariate Gaussian distribution (MVGD) and shape from shading techniques. The third step is performed with Basel face model to create 3D images.

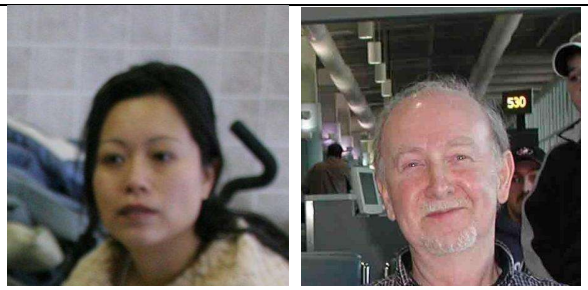
A. FEATURES EXTRACTION

Features extraction consists of two further steps. First, facial features such as nose, eyes, lips and mouth, etc. are extracted. Then SIFT is used to extract distinctive features.

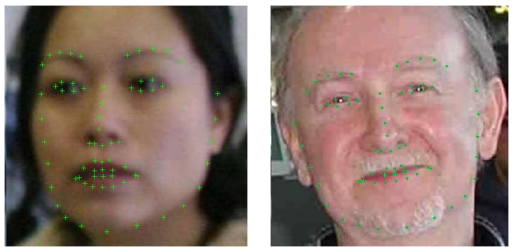
1) FACE DETECION AND FEATURES EXTRACTION

In this step, a single 2D image is taken as input. First, the image is scanned for face detection. If the input image has a face, it is further processed. Different landmarks are localized on the face. Images from LFW database [27] are being used for experimental purposes. The Viola-Jones method, which is further developed by a boosted frontal profile face detector, is used to detect faces [28]. Then, landmarks are estimated by using the method described in [29]. 68 landmarks are marked near the right and left eyes, nose, mouth, and boundary of chin & cheeks. Original images can be seen in Fig. 2a, and results for these landmarks can be seen in Fig. 2b. For a deeper understanding, reader can consult [29].

Images from the input are passed through several preprocessing steps that detect the face, mark the landmarks, and then crop the images according to their face resolution, so that it can be used for further processing.



a. Original images



b. Face detection and facial features extraction

FIGURE 2. Face Detection

2) DISTINCTIVE FEATURES EXTRACTION

The next step is to find distinctive features by using scale-invariant feature transform (SIFT) [30]. These features will help to reconstruct a 3D shape. Four stages are performed to find these features. First, difference of Gaussian (DOG) is used to detect potential interest points. These points are invariant to orientation and scale. This step is called scale space extrema detection. Next step is the localization of key points, where scale and location are determined for every candidate location. Then on local image gradient direction, each key point is assigned by one or more orientations. In the final step, around these key points, gradients of local images are measured.

One of the benefits of this approach is that it produces a number of dense features. These features cover the whole image: figure 3 describes some outputs of SIFT. Where green dots show the positions of feature points for the eyes, nose, and side of face for alignment purposes, the red squares enclose prominent green points that will help in aligning the image with a Basel model, which is described in section 3.

FIGURE 3. SIFT features or dense features extraction

B. DEPTH ESTIMATION

This section will estimate the depth of the input image. The distance between the sensor and the surface of a scene is called depth. It is an essential step to understand the 3D geometry. The estimation consists of two parts. First, depth will be calculated with the help of Multivariate Gaussian distribution (MVGD) [31]. Then, enhanced details will be calculated using the shape from shading technique. The depth will be calculated out of a single image using MVGD. An image has two dimensions before this step. By assigning the depth as the third dimension to the image, the image can be rotated and translated according to the need. By relating MVGD with the examined problem, its distribution function will be as follows, by letting vector 'a' of dimension 'S':

$$G_a[a \mid \mu, \sigma] = \frac{1}{(2\pi)^{S/2} \mid \sigma \mid^{1/2}} e^{1/2(a-\mu)^T \sigma^{-1}(a-\mu)} \quad (1)$$

Here μ is mean in the data and σ is covariance, which will be calculated by estimating the maximum depth likelihood.

$$\mu_{ML} = \frac{1}{n}\sum_{i=1}^{n} a_i \quad (2)$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} (a_i - \mu_{ML})(a_i - \mu_{ML})^T \quad (3)$$

After implementing it on image data, one problem arises. Since an image has a large number of rows and columns, multiplication makes it computationally complex, resulting in a lot of time spent. Principal component analysis (PCA) [30] is used to overcome this problem. PCA helps us to reduce the dimensions of matrix for multiplication. It finds a covariance matrix of data with the help of eigen decomposition. The first step of PCA calculation is to subtract the mean from the data.

$$I = [a_1 - \mu, ...a_n - \mu_n] \quad (4)$$

Covariance matrix σ can be found by singular value decomposition (SVD).

$$I = II^T \quad (5)$$

It helps to find the covariance matrix. However, as discussed above, multiplying these matrices will result in a high dimensionality matrix - which again makes it difficult for the computation. To overcome this problem, we multiply these matrices as shown in equation 6.

$$I = I^T I \quad (6)$$

Now the dimensions of matrix will be reduced, and calculations are simplified. Next step is to find depth by finding conditional distribution using the following equation

$$P(a_n / a_m) = G_a(\mu_n + \sigma_{nm}\sigma_{mm}^{-1}(a_m - \mu_m), \sigma_{nn} - \sigma_{nm}\sigma_{mm}^{-1}\sigma_{mn}) \quad (7)$$

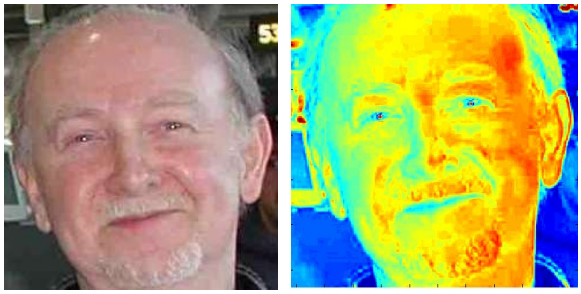
After calculating its distribution, weighted least squares (WLS) by Min et al. [33] are applied to find the depth, which is shown in equation 8.

$$D = \arg\min\{\sum_{i\in\Omega^0}(D_i - D_i^0)^2 + \alpha\sum_{i\in\Omega}\sum_{j\in N(i)}\omega_{i,j}^c(D_i - D_j)^2\} \quad (8)$$

Here D^0 are the measurements that are mapped as the depth, N(i) is neighborhood of I for the square patch - which can be set to desired size, and $\omega_{i,j}^c$ is defined as

$$\omega_{i,j}^c = \exp(\frac{|i-j|^2}{2\alpha_s^2})\exp(\frac{\sum_{k\varepsilon c}|I_i^k - I_j^k|^2}{3\times 2\alpha_s^2}) \quad (9)$$

Where C = {R, G, B} or C = {Y, U, V}, that represents different channels of the guidance color image. Depth of input image can be seen in Fig. 4. Colors in the figure of the depth map specify the estimated distances from the camera: yellow shows the nearest point to camera, while blue is the farthest point from camera.



a. Original Image b. Depth Estimation

FIGURE 4. Depth Estimation

At this stage, shape from shading technique (SFS) will be implemented to find enhanced details. SFS is a method which computes the 3D shape from one image of that shape. The main equation for image intensity is $I = \delta\lambda(N.L)$, which help us to find image intensity. Where λ is flux emission, δ is average reflectance, N is normal of surface, and L is illumination direction.

To find these details, surface of image is considered as a Lambertian surface. It is ideally considered to be a rough or matte surface, which isotopically scatters the incandescent light with Lambertian surface. The reflection function (Eq. 10) of Pentland is used for reconstruction of the surface.

$$I_{(x,y)} = \frac{p\cos\tau\sin\alpha + q\sin\tau\sin\alpha + \cos\alpha}{(p^2 + q^2 + 1)^{1/2}} \quad (10)$$

Here 'p' and 'q' are the slopes of the surface (Eq. 11), τ is tilt of illuminant (tilt is an angle that is made by the x-axis with image plane component of illuminant vector), and α is slant (the angle the z-axis makes with illuminant vector)

$$p = \frac{\partial}{\partial x}z(x,y)$$
$$q = \frac{\partial}{\partial y}z(x,y) \quad (11)$$

Now, equation 8 must be converted to such a form that is capable of helping us relate it with a 3D surface; so, Pentland takes the Taylor series and expands its reflectance function. It propagates the information along the shaded surface, which is starting from points with unknown orientation of surface and tries to estimate the shape from local variations in the intensity of the image.

C. CREATION OF 3D IMAGE

After implementing the above steps, we have data in three directions, so we can align our data with a 3D Morphable model. A Basel face model [34] that is publicly available (faces.cs.unibas.ch) was used. It consists of 100 females and 100 males' 3D scans. Now, we have to establish the point-to-point correspondence with BFM. We have data in three dimensions and also BFM. To fit this model, our main goal is to minimize the sum of square differences of pixels between the processed image and the given model. The sum can be written as

$$E = \sum_{x,y}\| I_{input}(x,y) - I_{model}(x,y)\|^2 \quad (12)$$

The 3D space face model of Blanz et al. [35] for alignment purposes is used. Its cost function is

$$E = \frac{1}{\sigma_I^2}E_1 + \frac{1}{\sigma_F^2}E_F + \sum_i\frac{\alpha_i^2}{\sigma_{T,i}^2} + \sum_i\frac{\beta_i^2}{\sigma_{T,i}^2} + \sum_i\frac{(\rho_i - \bar{\rho}_i)^2}{\sigma_{R,i}^2} \quad (13)$$

Here σ_I is standard deviation, which is the likelihood for I_{input}, and $\alpha\beta\rho$ is one dimensional product of normal distribution. $\beta = (\beta_1, \beta_1, ...)^T$ are texture coefficients, and shape coefficients are $\alpha = (\alpha_1, \alpha_1, ...)^T$. In the above equation, 13 is minimized by a stochastic newton algorithm, which is derived in [35]. So, after minimization of equation 13 we get

$$E_{I,approx} = \sum_k\left\|I_{input}(p_{x,k}, p_{y,k}) - I_{model,k}\right\|^2 \quad (14)$$

Here $(p_{x,k}, p_{y,k})$ defines the image-plane position. This equation helps to align an image on a given model.

IV. Evaluations

Evaluations are performed using different type of images, including ones sourced from the Labeled Faces in the Wild Home (LFW) database [27], online randomly searched celebrity images, and images used in [36] and [37]. An online image database, LFW is available for experimental purposes. This database has hundreds of images. So, some random images were taken to validate our proposed method. The proposed method is examined on a variety of images with different ages and variations. Qualitative and quantitative comparisons are performed with other latest methods which are shown below. Some randomly selected images, inputted as a single 2D image, were reconstructed into 3D using the proposed method. Figure 5 shows the results. Facial features, which are briefed in section 2, are shown in Fig. 5 (a); dense features, or SIFT features, are shown in (b); (c) and (d) show the 3D reconstructed face without texture; (e) and (f) display the textured 3D faces with two different views.

Comparisons. The experimental results show the satisfactory results for 3D reconstruction from a single 2D image. A performance comparison between the proposed method and some existing methods, both qualitatively and quantitatively, is conducted. For qualitative analysis, [36] and [37] are compared with our method, whereas quantitative analysis is performed by comparing our method with [38] and [39]. For qualitative comparison, internet celebrity pictures (George Clooney, Kevin Spacey, and Tom Hanks) are used, whereas quantitative comparison is performed on the images used in these papers (face I, face II and bs000). Qualitative comparison of the proposed method is displayed in Figs. 6 and 7.

Several images are randomly examined to check the validity of the proposed method. In Figures 6 and 7, (a) shows the reference images and (b) displays the results of the 3D face reconstructed by the existing methods [36] and [37], whereas (c) and (d) show the results of the proposed method. The figures show that our method is better than the existing methods. It can be seen that the 3D reconstruction of our method is more similar to the original image.

FIGURE 5. Reconstruction results of the proposed method. (a) the results after features extraction, (b) SIFT features, (c) 3D reconstructed image before texture, (d) side view of the reconstruction, (e) and (f) texture fitted faces with different poses.

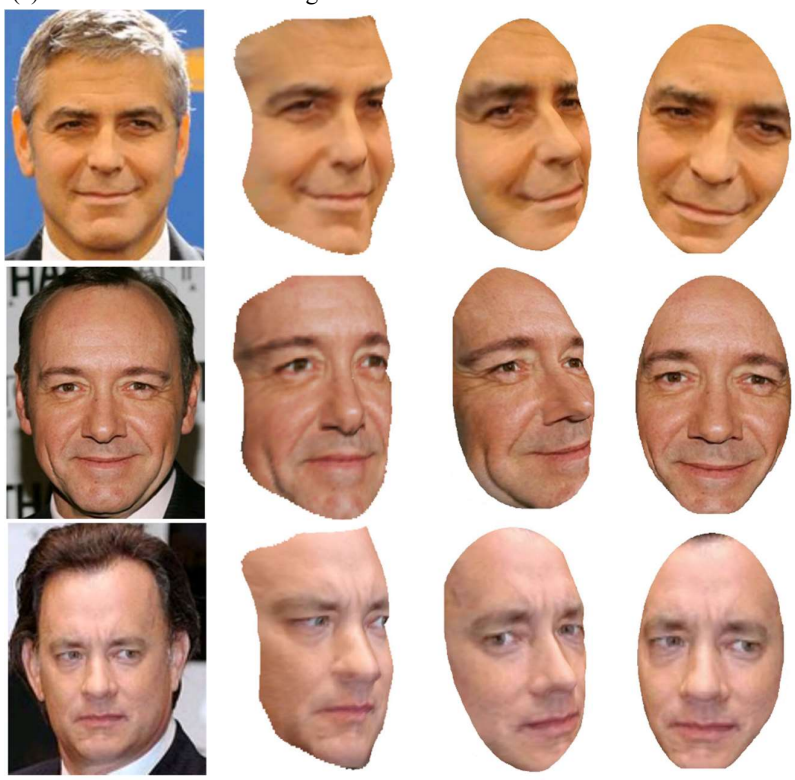
For example, the nose of images reconstructed by our method is sharper, and it more closely resembles the reference image compared to the output of the examined methods. The superiority of our method to the examined methods can also be found by checking wrinkles and dimples in the figures. Different quantitative values of these images are shown in table 1.

TABLE 1. ERROR CALCULATIONS OF RECONSTRUCTED IMAGES OF FIG. 6 & 7. (BETWEEN A & C)

Factors	George Clooney	Kevin Spacey	Tom Hanks	Face I [24]
RMSE	0.8866	0.8689	1.6468	0.8560
PSNR	27.275	29.433	28.523	35.626
Mean Error	0.4036	0.3790	0.3821	0.2043
Std. Error	0.3277	0.3487	0.3348	0.4163
SSIM	0.9906	0.9881	0.9829	0.9941

The table shows the different types of error calculations, which demonstrates the accuracy of our method. Values are calculated between the ground truth (a) and the reconstructed image from proposed method (c) in figures 6 and 7. Root mean square error (RMSE) is very low, and peak to signal noise ratio value is also very good. SSIM is structural similarity index for measuring image quality. From the results, it is clear that the image similarity index is much higher, which depicts that the reconstructed image optimally resembles the reference images.

The qualitative comparison between our method and other methods has demonstrated the superiority of our method. To consider the objective aspects, we have also compared our method quantitatively with more methods explained in [38] and [39]. For quantitative measurements, the following factors are calculated: root mean square error (RMSE), mean error, standard deviation (Std.) error, and structural similarity index SSIM. RMSE measures how much error is present between two data sets. Quantitative results are shown in tables 2 and 3. Comparison of the reconstructed images with [38] is displayed in table 2. [38] uses face I, face II and face III, which are shown in Fig. 8. One more method [39] is compared in table 3. Images used in [39] are shown in Fig. 9.



a b c d

FIGURE 6. Comparison of results with [36]. (a) The reference images from [36]. (b) 3D reconstructed images from [36]. (c) and (d) 3D reconstructed images of the proposed method.

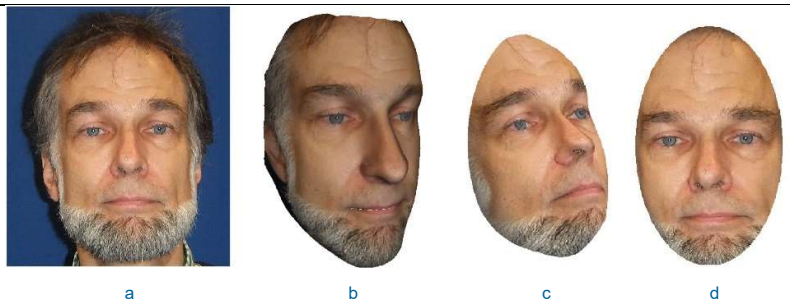


FIGURE 7. One more benchmark is presented. (a) The reference image taken from [37]. (b) The image reconstructed by [37]. (c) and (d) 3D reconstructed images of the proposed method.

FIGURE 8. Comparison of images used in [38]

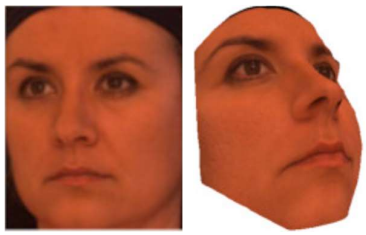


FIGURE 9. Images taken from [39] for comparison

TABLE 2. COMPARISON OF METHOD [38] WITH THE PROPOSED METHOD (RMS ERROR)

Methods	Face I	Face II	Face III
[38]	2.873	2.013	2.174
The proposed	0.856	0.842	0.849

TABLE 3. COMPARISON OF METHOD [39] WITH THE PROPOSED METHOD

Methods	SSIM	Error (%)	M. Err. (%)	Std. Err. (%)
[39]	0.9696	3.04	1.6468	0.8560
The proposed	0.9864	1.36	0.499	0.3070

Another major advantage of our method is that it doesn't need manual marking of landmarks as other existing methods do. As we have discussed in section 1, most existing algorithms need manual marking of landmarks, which can be time consuming and hectic. Sometimes, the user is inadequately trained to handle these landmarks; computational skills are required to mark these landmarks, which can make it difficult for beginners or those with only basic computer skills. Manual marking wastes a lot of time for implementation. Whereas our method can fully automatically generate 3D faces. So, a novice user who doesn't have computing knowledge can simply input the desired 2D image and get a 3D reconstructed image in return: there is no special need of computational skills. Our method gives liberty to novice and expert users alike. Several existing methods need user interference, which is difficult for novice users. On the contrary, our method is fully automatic and doesn't need any user interference.

In the future, the proposed method will be improved by expanding the database of images; it will include images with expressions as well and will be developed for facial-recognition purposes.

V. CONCLUSION

This paper proposes an automatic method which takes a single image at input and reconstructs a 3D image at output. It consists of three main steps. First, facial features and dense features are extracted. After finding the facial features, dense features are extracted by using the SIFT technique. Then, its depth is found by using multivariate Gaussian distribution. Shape from shading technique, which works on Lambertian reflectance law, is performed to find its high-resolution details. By implementing shape from shading technique, we can recover high details - like dimples and wrinkles - efficiently. Our results section depicts that the proposed method recovers wrinkles efficiently. At the end, our processed data is aligned with a publicly available 3D face model, which is known as a Basel face model. The proposed method is validated by performing experimentations and comparisons with existing methods. Qualitative and quantitative results are compared with some advanced methods to show the accuracy of this method. By examining RMS error, it can be seen that the proposed method is efficient in nature. Results show that this method produces good results and is capable of reconstructing a 3D image with satisfactory results.

REFERENCES

[1] J. Luo, J. Zhang, B. Deng, H. Li, and L. Liu. "3D face reconstruction with geometry details from a single image." IEEE Transactions on Image Processing, vol. 27, no. 10, pp 4756-4770, 2018.

[2] X. Tu , J. Zhao, Z. Jiang, Y. Luo, M. Xie, Y. Zhao, L. He, Z. Ma, and J. Feng. "Joint 3D Face Reconstruction and Dense Face Alignment from A Single Image with 2D-Assisted Self-Supervised Learning." arXiv preprint arXiv:1903.09359, 2019.

[3] C. Ding, and T. Dacheng, "A comprehensive survey on pose-invariant face recognition." ACM Transactions on intelligent systems and technology (TIST), vol 7, no. 3, 2016.

[4] I. A. Kakadiaris, et al. "3D-2D face recognition with pose and illumination normalization." Computer Vision and Image Understanding, pp 137-151, 2017.

[5] J. F. Hu, W. S. Zheng, X. Xie, and J. Lai, "Sparse transfer for facial shape-from-shading." Pattern Recognition, vol. 68, pp. 272-285, 2017.

[6] I. Katsushi. "Lambertian Reflectance." Encyclopedia of Computer Vision. Springer. pp. 441–443, 2014.

[7] Y. Quéau, J. Mélou, F. Castan, D. Cremers, and J. D Durou, "A variational approach to shape-from-shading under natural illumination." In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 342-357, 2017.

[8] A. L. Jeni, F. C. Jeffrey and K. Takeo, "Dense 3D face alignment from 2D video for real-time use." Image and Vision Computing , pp 13-24, 2017.

[9] A. Maghari, I. Liao, and B. Belaton, "Quantitative analysis on PCA-based statistical 3D face shape modelin." Computational Modelling of Objects Represented in Images III: Fundamentals, Methods and Applications, vol. 13, 2012.

[10] S.F Wang, and S.H. Lai, "Reconstructing 3D face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp. 2115–2121, 2011.

[11] L. Zhang, N. Snavely, N. Curless, and S. M. Seitz, "Spacetime faces: high resolution capture for modeling and animation." ACM Trans. Graphics (Proc. SIGGRAPH), pp. 548–558, 2004.

[12] S. G. Gunanto, H. Mochamad and Y. M. Eko, "Feature-points nearest neighbor clustering on 3D face models." Cyber and IT Service Management, International Conference on. IEEE, pp 1-5, 2016.

[13] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, "High-quality single-shot capture of facial geometry." ACM Transactions on Graphics (TOG), Vol 29, 2010.

[14] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, Robert, W. Sumner, and M. Gross, "High-quality passive facial performance capture using anchor frames." ACM Transactions on Graphics (TOG), vol. 30, pp 75, 2011.

[15] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer, "High resolution passive facial performance capture." ACM Transactions on Graphics (TOG), vol. 29, 2010.

[16] J. Jo, H. Choi, I. J. Kim, and J. Kim, "Single-view-based 3D facial reconstruction method robust against pose variations." Pattern Recognition, vol. 48, no. 1, pp. 73-85, 2015.

[17] I. K. Shlizerman and S. M. Seitz. "Face Reconstruction in the Wild." International Conference on Computer Vision (ICCV), Nov 2011.

[18] A. Jourabloo and X. Liu. "Large-pose face alignment via cnn-based dense 3d model fitting." In the proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

[19] Y. Li, H. Su, C. R. Qi, N. Fish, D. Cohen-Or, and L. J. Guibas. "Joint embeddings of shapes and images via cnn image purification." ACM Trans. Graph, 5, 2015.

[20] H. M. R. Afzal, S. Luo, and M. K. Afzal, "Reconstruction of 3D facial image using a single 2D image." In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-5, March 2018.

[21] K. Jiang, Z Wang and P. Yi, "ATMFN: Adaptive-threshold-based Multi-model Fusion Network for Compressed Face Hallucination." IEEE Transactions on Multimedia, vol. 1 no. 1, 2019.

[22] F. Liu, Q. Zhao, X. Liu and D. Zeng, "Joint Face Alignment and 3D Face Reconstruction with Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 664-678, 1 March 2020.

[23] Y. Zhou, W. Zhang, X. Tang, and H. Shum. "A bayesian mixture model for multi-view face alignment." In proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 1741–1746, 2005.

[24] 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, Jul. 2005.

[25] J. Chen, V. M. Patel, L. Liu, V. Kellokumpu, G. Zhao, M. Pietikäinen, and R. Chellappa, "Robust local features for remote face recognition." Image and Vision Computing, vol. 64, pp. 34-46, 2017.

[26] C. Cao, Y. Weng, S. Lin, and K. Zhou, "3d shape regression for real-time facial animation." ACM Trans. Graph. 32, 4, July, 2013.

[27] B. G. Huang, M. Ramesh, and T. Berg,. "Labeled Faces in the Wild: A Database for Studying Face Recognition" in Unconstrained Environments, 2008.

[28] Z. Kalal, J. Matas, and K. Mikolajczyk. "Weighted sampling for large-scale boosting." In BMVC, pp. 1-10, 2008.

[29] X. Zhu, and D. Ramanan. "Face detection, pose estimation and landmark localization in the wild." Computer Vision and Pattern Recognition (CVPR) Providence, Rhode Island, pp. 2879-2886, June 2012.

[30] S. Chen, S. Zhong, B. Xue, X. Li, L. Zhao, and C. I. Chang, "Iterative Scale-Invariant Feature Transform for Remote Sensing Image Registration." IEEE Transactions on Geoscience and Remote Sensing, 2020.

[31] C. Grana, G. Serra, M. fredi, and R Cucchiara, "Image Classification with Multivariate Gaussian Descriptors. In: Petrosino A. (eds) Image Analysis and Processing." Lecture Notes in Computer Science, vol 8157. Springer, Berlin, Heidelberg. 2013.

[32] A. Maghari, I. Liao, and B. Belaton, "Quantitative analysis on PCA-based statistical 3D face shape modeling." Computational Modelling of Objects Represented in Images III: Fundamentals, Methods and Applications, vol. 13, 2012.

[33] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares." IEEE Trans. Image Process, vol. 23, no. 12, pp. 5638–5653, 2014.

[34] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D Face Model for Pose and Illumination Invariant Face Recognition." In Proceedings of the 6th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments Genova (Italy), pp 296 - 301, Sep. 2009.

[35] Blanz, and T. Vetter, "Face recognition based on _tting a 3D morphable model." IEEE Trans. Pattern Anal. Mach. Intell. 25, pp 1063-1074, 2003.

[36] I. K. Shlizerman and S. M. Seitz. "Face Reconstruction in the Wild." International Conference on Computer Vision (ICCV), Nov 2011.

[37] H. Jain, O. Hellwich, and R. S. Anand. "Improving 3D face geometry by adapting reconstruction from stereo image pair to generic Morphable Model." Information Fusion (FUSION), 19th International Conference on. IEEE, pp. 1720-1727, 2016.

[38] H. Jain, O. Hellwich, and R. S. Anand. "Improving 3D face geometry by adapting reconstruction from stereo image pair to generic Morphable Model." Information Fusion (FUSION), 19th International Conference on. IEEE, pp. 1720-1727, 2016.

[39] W. Peng, C. Xu, and Z. Feng. "3D face modeling based on structure optimization and surface reconstruction with B-Spline." Neurocomputing, pp. 228-237, 2016.