Reconstruction of 3D facial image using a single 2D image

H. M. Rehan Afzal, Suhuai Luo
School of Electrical Engineering and Computing
University of Newcastle, Newcastle, Australia
nwpu_edu_cn@yahoo.com

M. Kamran Afzal
School of Computer Science and Technology
Xiamen University
Xiamen, China

***Abstract*— In many scenarios related to image processing, 3D face reconstruction is considered an essential part. A robust and fast method of 3D face reconstruction from a single 2D image is presented. This method consists of three main steps including extraction of features from single image, calculating the depth of image and adjustment of a 3D model on the direction of Z-axis. In the first step, the features of image are extracted by using supervised descent method (SDM). Using SDM, face regions like facial components (eyes, nose lips) and face contours are detected. Second step consists of depth prediction by implementation of multivariate Gaussian distribution. Finally, 3D face is constructed with the help of features and the depth information and 3D database. The proposed method has been verified by conducting several experiments depicted in evaluation section. Our method is robust in nature and gives good results even using a single image, comparing to other methods that use multiple images for reconstruction of 3D images.**

Keywords—3D face reconstruction; features extraction; Gaussain distribution; facial modeling

I. INTRODUCTION

Due to complex structure, three dimensional (3D) object reconstruction is always very difficult and a challenging task. Image processing engineers also face such problems while handling 3D environment. 3D face reconstruction has always fascinated attention because of its wide applications [1]. It has numerous applications in almost every filed, like medical, animations, computer games, robotics and security etc. Past two decades show the significance of face reconstruction due to increment of computer aided programs. Recently huge research is being carried out on 3D face reconstruction for medical images and face recognition. 3D images help us to increase accuracy and precision for further processing. A number of methods [2-7] are designed to construct 3D images by researchers in image processing.

Mostly methods need multiple images for reconstruction which is time consuming and computationally expensive. Researchers are paying attention to reconstruct 3D image from single image because in practical applications, it has high value. So, image reconstruction from single image has been implemented by researchers in past years which is called shape from X. First attempt for this approach was done by Horn [8] which is called shape from shading technique. He proposed human face as Lambertian surface which depicts that face follows the Lambert's law. Then he found depth by calculating reflectance function using Taylor series and used first order approximations. Later, Pentland [9] also implemented shape from shading techniques and calculated surface albedo and depth by implementing fast fourier transformation (FFT) on Lambertian reflectance function. Tsai & Shah [10] calculated reflectance function by employing liner approximation. For surface normal, they used discrete approximation with help of finite differences. These methods (shape from shading) were fast in nature but considered ill posed due to several reasons. Firstly, the depth information was hard to find from image. Secondly, all images don't follow Lambertian reflectance law, so specular surfaces were very hard to reconstruct. When synthetic data is given, these shape from shading algorithms give poor results. For real images, results are even worse.

After that, learning based approaches [11], [12], [13] were proposed by researchers. These approaches learn the information from the data which is called training data for 3D reconstruction. A well-known approach was proposed by Blanz and Vetter [14]. By Principal factor analysis, a statistic 3D Morphable Model (3DMM) was calculated by them. Then this statistic model was trimmed with 2D image for the reconstruction of 3D face. Later a lot of approaches followed this technique and generated good results. Some regression approaches also taken part in this competition. These regression techniques explore the distributions for the training faces and regression was calculated between them. Training data helped to learn the useful information. One more type of these models was cascaded regression. It simply combines strong regressor into weak regressors. These methods have good results for specific scenarios and still have limitations. As we discussed above that shape from shading methods works for those surfaces which obeys Lambertian reflectance law, so there is always need of a new method.

In this paper, a 3D face reconstruction is proposed. A single 2D image is taken at input and 3D image is constructed at output. Proposed method is depicted in section II. Results are evaluated in section III whereas section IV gives the conclusion of proposed method.

978-1-5386-1370-2/18/$31.00 ©2018 IEEE

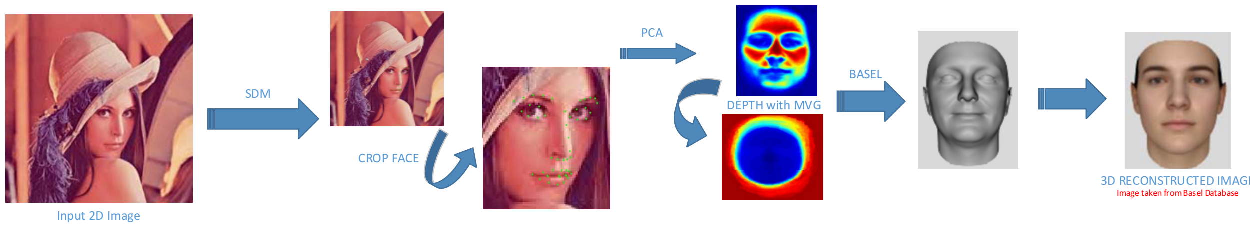


Fig. 1. General overview of proposed method

II. The Proposed Method

In this section, a new method is proposed which will take a single 2D image and will reconstruct it into 3D image. Detailed general diagram of proposed method can be seen in Fig. 1.

This method is mainly consisting of three parts. First of all, facial features such as mouth, nose and eyes (left and right eye) are detected using Supervised Descent Method [15] (SDM) and face is detected according to these facial features. Second step consists of prediction of depth calculation with the help of Multivariate Gaussian Distribution (MVG). At the end third step comes with 3D face reconstruction. The features calculated in first step and depth calculation help us to reconstruct 3D face according to a 3D database. Basel Face Model (BFD) is being used as a 3D database. These three steps are discussed briefly in the following sub sections.

A. Face Detection & Features Extraction

First stage of our method is to find the required features. For this purpose, a method named SDM is used which has good results and robust in nature. It works simply by minimizing a function which is called nonlinear least square function. The main benefit of this algorithm is that while minimizing nonlinear square function, it uses learned decent direction which don't need Hessian or Jacobian for computations process. It decreases a lot of computations and makes system fast.

Before facial detection, face detector is implemented on the training image which helps us to find initial arrangement of features. With the help of face detector, we get two benefits. 1) It will help us to check whether an image has a face or not. If an image doesn't has a face then it will not go for further processing. Face detector has function which help us to differentiate the image with and without face. When our method confirms that examined image has a face, it gives positive alarm and go to further step. 2) It will assist SDM for better landmark detection. Let we have an image I with m pixels, h is a nonlinear function for feature extraction. So initial framing for face alignment can be minimized over Δy .

$$f(y_0+\Delta y)=\left\|h(I(y_0+\Delta y))-\varphi_*\right\|_2^2 \qquad (1)$$

Here y_0 is landmark, $\varphi_* = \mathrm{h}(I(y_0))$. After applying Taylor series and taking derivative on Eq. 1 with respect to y_0, we get following equation.

$$\vartriangle y_1 = R_0\varphi_0 + b_0 \qquad (2)$$

Here $R_0=-2H^{-1}J_h^T$ and $b_0=(\varphi_0-\varphi_*)$ which give us first update for x.

SDM was tested on LFW Dataset [16] which is available online. Fig. 2 depicts the results of our first step of method. First, face is detected for input image, then features are extracted.

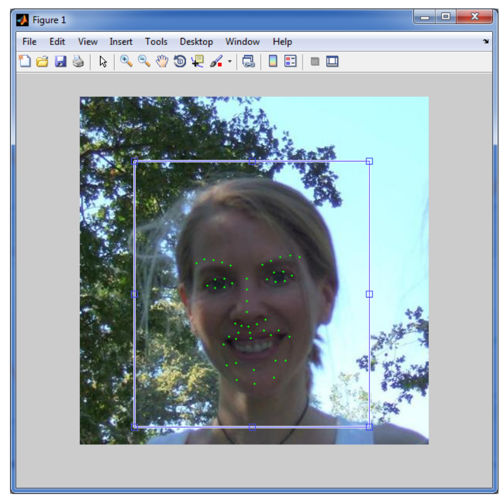


Fig. 2. Feature selection and face detection

Fig. 2. Gives us features of face (eyes, nose and mouth). Then we crop this image and calculate depth which is mentioned in next section. Some more images after features extraction and cropping can be seen in Fig. 3.

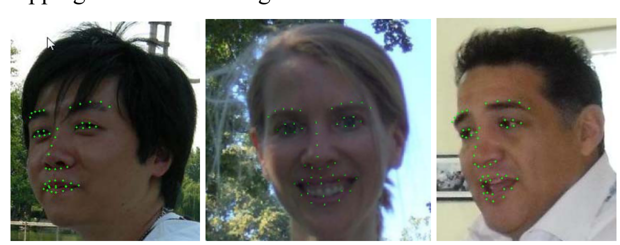
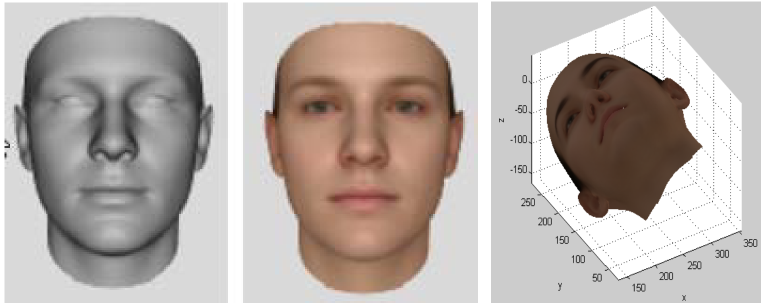


Fig. 3. Feature extraction and cropping



BFM mean shape BFM mean texture Proposed method generated 3D Image

Fig. 5. Basel Face Model and Proposed method image

B. Depth Prediction using Multivarite gaussain Distribution (MVG)

Now we have extracted features. Next step is to find the depth of image to find third dimension or z dimension. We take previous dataset found in above section and then model it by applying Multivariate Gaussian Distribution [17].

Let we have vector ‘a’ of dimension ‘S’, so according to MVG our distribution function will be

$$G_a[a \mid \mu, \sigma] = \frac{1}{(2\pi)^{S/2} \mid \sigma \mid^{1/2}} e^{1/2(a-\mu)^T \sigma^{-1}(a-\mu)} \tag{3}$$

Here μ is mean in the data and σ is covariance. These two can be calculated by estimating the maximum likelihood

$$\mu = \frac{1}{n} \sum_{i=1}^{n} a_i \tag{4}$$

$$\sigma = \frac{1}{n} \sum_{i=1}^{n} (a_i - \mu_{ML})(a_i - \mu_{ML})^T \tag{5}$$

Depth calculation can be seen in Fig. 4.

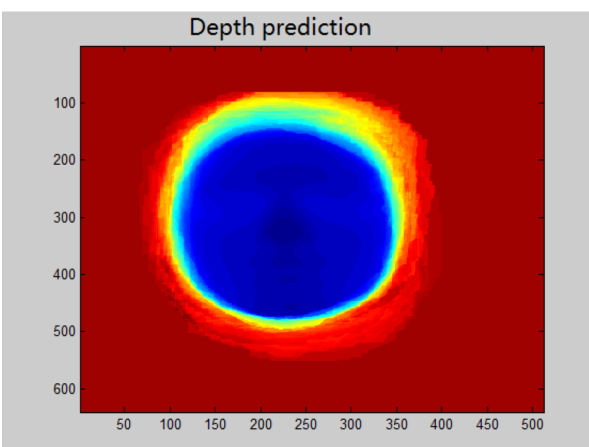


Fig. 4. Depth Prediction

III. 3D Face Allignment with 3D Database

Our final step is to align positions of 2D landmarks with 3D Morphable model. For this purpose, we have used Basel face model [20]. It consists of 100 female and 100 male 3D scans. It is publically available (http://faces.cs.unibas.ch/).

At this stage, we have facial features in 2D domain and we also calculated depth of input image. So, we can rotate image according to requirement. It is easy for us to align our image with BFM. Our main purpose is to establish point to point correspondence with above data and BFM. We use Blanz et al. [14] 3D face space model for alignment purposes. A cost function is proposed.

$$E = \frac{1}{\sigma_I^2} E_I + \frac{1}{\sigma_F^2} E_F + \sum_i \frac{\alpha_i^2}{\sigma_{T,i}^2} + \sum_i \frac{\beta_i^2}{\sigma_{T,i}^2} + \sum_i \frac{(\rho_i - \bar{\rho}_i)^2}{\sigma_{R,i}^2} \tag{10}$$

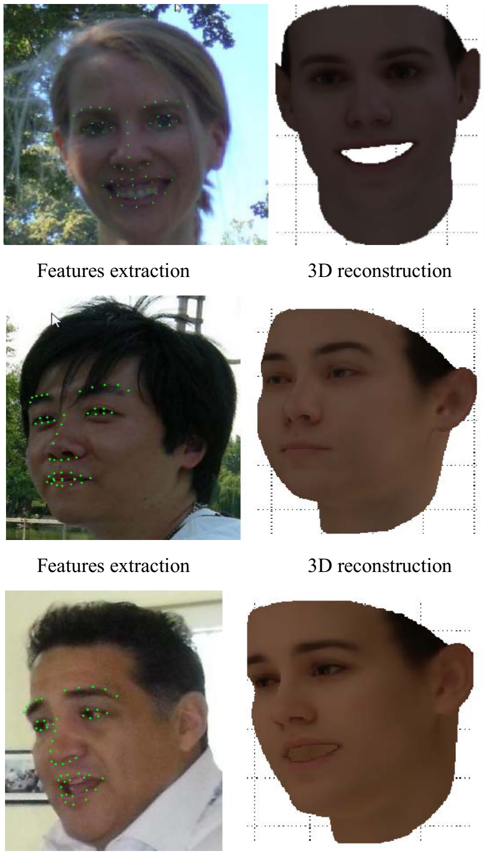
Here σ_I is a standard deviation, which is likelihood for I_{input} and $\alpha\beta\rho$ is one dimensional product of normal distribution. $\beta = (\beta_1, \beta_1, ...)^T$ are texture coefficients and shape coefficients are $\alpha = (\alpha_1, \alpha_1, ...)^T$ and other rendering parameters like pose angles ϕ, θ ,and γ, focal length f, translation t_w , direct light intensities $L_{r,dir}, L_{g,dir}, L_{b,dir}$, ambient light intensities $L_{r,amb}, L_{g,amb}, L_{b,amb}$ and offset of color channels $g_r, g_g, g_b, o_r, o_g, o_b$. After minimization Eq. 10 we get

$$E_{I,approx} = \sum_k \left\| I_{input}(p_{x,k}, p_{y,k}) - I_{\mathrm{model},k} \right\|^2 \tag{11}$$

A. Evaluations

Results were taken by implementing proposed method on images taken from online database LFW [16].

A portion of images were selected from this database and examined with proposed method. Fig. 5 shows the BFM mean shape, mean texture and proposed method image. After implementing our method with the help of BFM we present results in Fig. 6.



Features extraction 3D reconstruction

Features extraction 3D reconstruction

Features extraction 3D reconstruction

Fig. 6. Proposed method results

Comparison

The above proposed method has several benefits as compared to existing methods. If we compare it with existing methods which take multiple images at input, our method is fast for computations. These methods take multiple images at input and learn data after huge calculations but in our case our method simply takes one single image which reduces calculations and makes our system fast.

Our second comparison is with shape from shading (SFS) 3D reconstruction techniques. SFS techniques work well when image follows Lambertian reflectance law but fails badly for specular images. It also gives bad results for real images. Whereas our method works well for every type of images. From results it can be seen that it works well for real images. The 3D reconstructed images of this method have good results. Our method is scalable, modular and flexible. The proposed method also works fine for different facial expressions, as shown in Fig. 6. Where, images of row 1 and row 3 have open mouth expressions. It not only worked for frontal images but also reconstructed images efficiently for side posed images. One more purpose of 3D face reconstruction is that it can help us for efficient face recognition. Experiments show that 3D reconstructed faces give good results for face recognition as compared to 2D face algorithms. In future this work will be expanded for a face recognition algorithm. Our method gives liberty to novice users as well as expert users. Some existing methods need user interference which is difficult for novice users. On contrary, our method is fully automatic and don't need any user interference which increases its worth.

IV. Conclusion

In this paper, we have proposed a method which takes a single 2D image as input and outputs a 3D reconstructed image. This method is fast and robust. It consists of three steps. First step consists on feature extraction. For feature extraction we have used SDM which is computationally efficient. Face is detected at first, then facial features like nose, eyes and mouth are extracted. This image was in 2D dimension, so we applied multivariate Gaussian distribution to find its depth which added one more dimension in our examined image. To overcome the problem of huge computations, PCA was implemented to reduce dimensions.

Finally, the data gained from above two steps were aligned with 3D Basel face model to reconstruct 3D face. Result section shows the final outcome of 3D face reconstruction. This method had been tested on a number of images taken from LFW database and its accuracy was checked. Most of previous algorithms used multiple images as input to train algorithms which are computationally costly. However, the proposed method only takes a single 2D image and efficiently reconstructs 3D image.

References

[1] Aldrian, O., Smith, W.A.P.: Inverse rendering of faces with a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1080-1093.

[2] M. Castelan, W.A.P. Smith, and E.R. Hancock, "A Coupled Statistical Model for Face Shape Recovery from Brightness Images," IEEE Trans. Image Processing, vol. 16, no. 4, pp. 1139- 1151, 2007.

[3] J. Xie, R. Girshick, and A. Farhadi. Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks. In ECCV, 2016.

[4] I. Kemelmacher and R. Basri, "Molding Face Shapes by Example," Proc. European Conf. Computer Vision, vol. 1, pp. 277-288, 2006.

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

[6] Godard, C., Mac Aodha, O., & Brostow, G. J. (2016). Unsupervised monocular depth estimation with left-right consistency. arXiv preprint arXiv:1609.03677. Chicago

[7] X. Zhu, Z. Lei, J. Yan, D. Yi, S. Z. Li, "High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

[8] B.K.P. Horn, "Obtaining Shape from Shading Information," The Psychology of Computer Vision, McGraw-Hill, 1975.

[9] A.P. Pentland, "Finding the Illuminant Direction," J. Optical Soc. Of Am., vol. 72, pp. 448-455, 1982.

[10] P.S. Tsai and M. Shah, "Shape from Shading with Variable Albedo," Proc. Optical Eng., pp. 121-1220, Apr. 1998.

[11] Lin, C., Cheng, W., Liang, S.: Neural-network-based adaptive hybrid-reflectance model for 3D surface reconstruction. IEEE Transactions on Neural Networks 16(6), 1601–1615 (2005).

[12] Maghari, A., Liao, I., Belaton, B.: Quantitative analysis on PCA-based statistical 3D face shape modeling. In: Computational Modelling of

Objects Represented in Images III: Fundamentals, Methods and Applications, vol. 13 (2012).

[13] Wang, S.F., Lai, S.H.: 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 33(10), 2115–2121 (2011).

[14] Blanz, V., Vetter, T.: Face recognition based on _tting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1063-1074.

[15] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. CVPR, 2013.

[16] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.

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

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

[19] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[20] Paysan P.,Knothe R., Amberg B., Romdhani S., and Vetter T. 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, September, 2009.