IMPROVING 3D FACE RECONSTRUCTION FROM A SINGLE IMAGE USING HALF-FRONTAL FACE POSES

Maurício Pamplona Segundo, Luciano Silva, Olga Regina Pereira Bellon*

IMAGO Research Group, Universidade Federal do Paraná, Brazil

ABSTRACT

In this work we evaluate the influence of pose variation on 3D face reconstruction from a single image. To this end, we present a 3D reconstruction method that combines a fitting technique and a sparse 3D deformable model to estimate the 3D information of 2D images with large pose variations. For our experiments, we synthetically created 2D images by rendering 3D models from the BU-3DFE database in different points of view. Thus, we have a precise ground truth that allows performing a quantitative analysis of the reconstruction accuracy. Our experimental results show that the reconstruction achieves the highest accuracy when using half-frontal face images, and is also more robust to noise and incorrect facial landmarks positioning.

Index Terms— Reconstruction algorithms, face recognition, biometrics.

1. INTRODUCTION

Recognizing faces of non-cooperative subjects using 2D images is a challenging problem due to variations in pose, illumination and facial expressions. To overcome these problems, many recent works in the literature have employed 3D face reconstruction methods [1–7].

Some 3D face reconstruction methods use multiple 2D images to recover the geometry of a face. Choi *et al.* [2] used the sparse bundle adjustment algorithm over a set of facial landmarks in five images with specific facial pose to compute a sparse 3D face model. A dense 3D face model is obtained by Medioni *et al.* [6] by applying a structure from motion technique to a high resolution video sequence. However, in some cases, there is only a single 2D image available to recover the geometry of a face.

Nevertheless, it is still possible to three-dimensionally reconstruct a face using a single 2D image as input. Jiang *et al.* [3] fitted a 3D deformable face model to a set of landmarks in a frontal face image to obtain a 3D face model. Blanz and Vetter [1] also used a 3D deformable face model, but the fitting process was guided by the texture information and there was no restrictions on the face pose. Finally, Kemelmacher-Shlizerman and Basri [4] used a shape from shading technique to deform a reference face model and obtain the 3D model of a face. All these face reconstruction methods can be performed fully automatically. However, in case of failures or unexpected scenarios (*e.g.* high illumination variation or painted faces), only landmark-based methods can be manually assisted in a practical way.

In this work, we developed a landmark-based face reconstruction method to recover the geometry of a face using a single image as input. To this end, the Levenberg-Marquardt (LM) iterative minimization technique [8] is applied to obtain camera and deformation parameters that fit a sparse 3D deformable face model in a set of facial landmarks located in the input image. These landmarks may be automatically located by an active face model [9] or manually obtained. After that, the Thin-Plate Splines (TPS) technique [10] can be used to deform a generic face model in order to obtain a dense 3D model of the input face, as in [2, 11].

Half-frontal face images have not been applied to face recognition as much as frontal [12, 13] and profile face images [14, 15] due to the difficulty of standardizing acquisition and appearance of such images without any 3D information. However, as observed in neuropsychology works [16], half-frontal images allow perceiving all three axes in a single image, as shown in Fig. 1, which is an interesting property for 3D face reconstruction. To investigate this observation, we applied our face reconstruction method to a wide range of pose variation in order to evaluate the influence of pose on 3D face reconstruction.
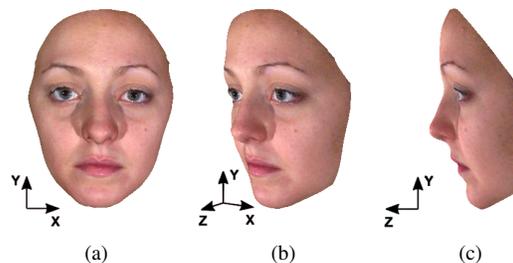
Fig. 1. Illustration of visible axes in face images with different pose: (a) frontal, (b) half-frontal, and (c) profile images.

In our experiments we used neutral face images of all subjects present in the BU-3DFE database [17]. Each face image is composed of a textured surface mesh and 83 facial landmarks, and was rendered in multiple poses to generate synthetic 2D images with a precise ground truth. Therefore, it was possible to quantitatively analyze the reconstruction accuracy and to compare the obtained results on renderings with different face poses.

2. 3D FACE RECONSTRUCTION

Our reconstruction method can be described as follows: given a set of 2D landmarks $P = \{p_1, p_2, \ldots, p_N\}$ in an input image, with $p_i = \{x_i, y_i\}$, the objective is to find the set $Q = \{q_1, q_2, \ldots, q_N\}$ with the 3D coordinates of the landmarks $q_i = \{X_i, Y_i, Z_i\}$ and the transformation T that minimize Equation 1:

$$\frac{1}{N} \sum_{i=1}^{N} ||p_i - p_i'|| \tag{1}$$

where $p_i' = Tq_i$. The transformation T is given by a camera model with seven parameters: translation in all axes t, rotation in all axes R and focal length f. A 3D landmark q_i is transformed into 2D coordinates on the input image p_i' by applying Equations 2-4:

$$q_i' = Rq_i + t \tag{2}$$

$$x_i' = \frac{W}{2} - \frac{X_i'}{Z_i'} f \tag{3}$$

$$y_i' = \frac{H}{2} + \frac{Y_i'}{Z_i'} f \tag{4}$$

where W and H are the width and height of the input image respectively.

The set of 3D landmarks Q is obtained through a sparse 3D deformation model. This model is created by applying Principal Component Analysis (PCA) [18] to training images centered on the average set of 3D landmarks $\bar{Q}$. After that, Q is also represented as a vector of weights $[w_1, w_2, \ldots, w_K]$, where w_i is the weight of the i-th eigenvector u_i returned by PCA. The original representation of Q is recovered by Equation 5:

$$Q = \bar{Q} + \sum_{i=1}^{K} w_i u_i \tag{5}$$

Since the size of the vector of weights (*i.e.* K) is much less than the size of the original representation (*i.e.* $3N$), this alternative representation is used during the reconstruction process to reduce the number of parameters.

The LM iterative minimization technique [8] is performed to find camera and deformable model parameters that minimize Equation 1 in order to obtain the final set of 3D landmarks.

2.1. Enhanced reconstruction through face symmetry

As the pose variation increases, the visible area of a face decreases. This fact is due to the self-occlusion problem. Fig. 2 shows the average number of visible landmarks in different face poses by applying horizontal rotations (*i.e.* right to left rotation of the head). As may be seen, the number of visible landmarks decreases with increasing face rotation.
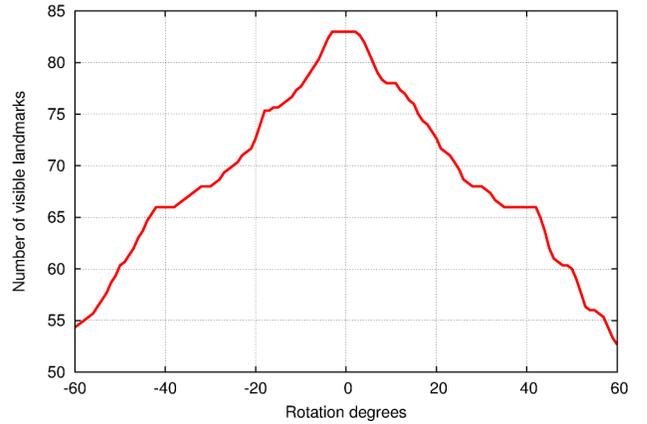
Fig. 2. Number of visible landmarks in different face poses.

However, most human faces are quite symmetric, which allows using the visible side of a face to estimate the occluded side. To this end, given two symmetric 3D landmarks q_i and q_j, q_j is mirrored into q_i to create a redundant 3D landmark $q_i^s = \{-X_j, Y_j, Z_j\}$. After that, the minimization equation in the parameter estimation stage is replaced by Equation 6:

$$\frac{1}{2N} \sum_{i=1}^{N} ||p_i - p_i'|| + ||p_i - p_i^s|| \tag{6}$$

where $p_i^s = Tq_i^s$. Besides providing valid information for occluded landmarks, this enhanced reconstruction method also reduces the influence of noise by providing redundant information for symmetric landmarks when both are visible.

3. EXPERIMENTAL RESULTS

Our experiments were designed to evaluate the accuracy of the presented reconstruction method across large pose variations. To this end, we used the available neutral images of the BU-3DFE database [17], totalling 100 images (one per subject). Each image has a textured surface mesh and a set of 83 facial landmarks that was used as ground truth in our experiments.

First we divided the database in training and testing sets, each one containing 50 images. After applying PCA to the training set, we keep 99% of the overall data variance.

After that, we used the testing set to create synthetic images with different poses, as illustrated in Fig. 3. For each testing image, 121 renderings were created with pose variation ranging from -60 to $+60$ degrees. Each rendering was reconstructed multiple times after adding a random noise with average magnitude ranging from 0 to 5 pixels to the facial landmarks' position in order to show the robustness of the method against noisy data.

(a)   (b)   (c)
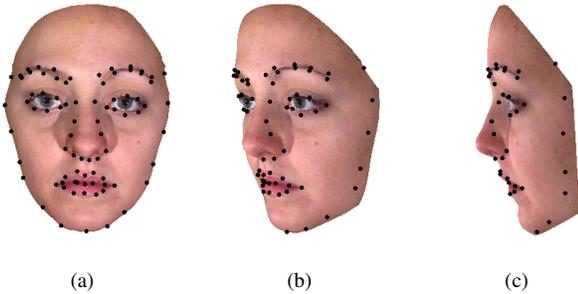
Fig. 3. Renderings of a BU-3DFE subject in (a) frontal, (b) half-frontal and (c) profile poses.

In our experiments, the reconstruction error is the average Euclidean distance between the ground truth and the final set of 3D landmarks in millimeters.

3.1. Reconstruction results

In our first reconstruction experiment, only frontal synthetic images were used. These images were reconstructed with the original number of landmarks, and then we simulated self-occlusion by removing visible landmarks, as illustrated in Fig. 4, in order to evaluate the robustness of the reconstruction method against missing information.

(a)   (b)   (c)
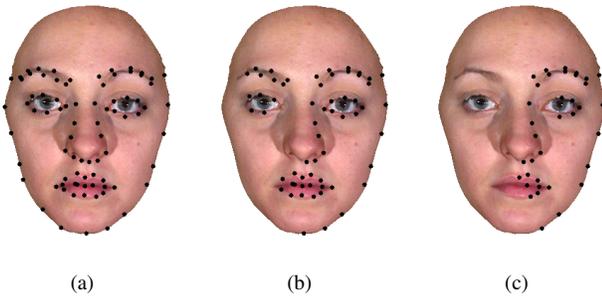
Fig. 4. Self-occlusion simulation in a frontal face image: (a) original image, and occlusion from (b) half-frontal and (c) profile images.

The obtained results are shown in Fig. 5, where the curve with label "3D" presents the average reconstruction error of

the ground truth, and other curves with label "2D / X px" present the average reconstruction error of frontal synthetic images with self-occlusion simulation, where X is the average noise magnitude in pixels. As may be seen, our sparse 3D deformable model is not able to exactly represent the images of the testing set since there are no subjects in both training and testing sets. For this reason, we obtained an average reconstruction error of about $2mm$ for the ground truth. Also, in all cases, the average reconstruction error increases with increasing self-occlusion and noise, as expected.
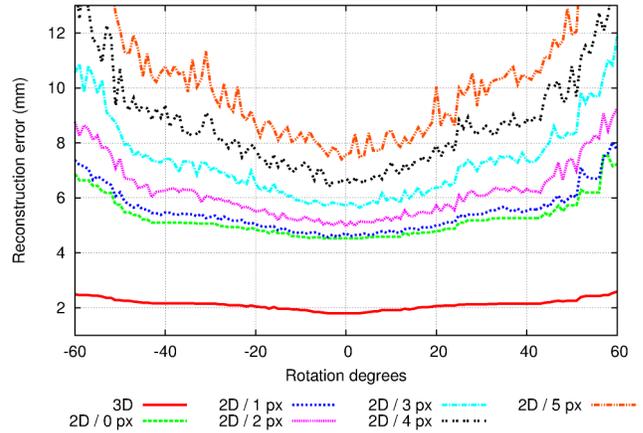
Fig. 5. Average reconstruction error of frontal synthetic images with self-occlusion simulation.

All synthetic images with various poses were used in our second experiment in order to evaluate the influence of pose variation on the reconstruction results. In the obtained results, shown in Fig. 6, the reconstruction error remains approximately constant in a wide range of rotation (*i.e.* from -45 to $+45$ degrees). Also, there is a slight advantage in using half-frontal images for low noise images. These results clearly outperformed the results shown in Fig. 5, evidencing the superiority of half-frontal images over frontal images, as observed in Fig. 1. However, they are still being affected by noise and large pose variations.

In our final experiment, we recompute the previous experiment using the enhanced reconstruction method (Sec. 2.1), and the obtained results are shown in Fig. 7. We have achieved high reconstruction accuracy and robustness against large pose variations and noise. The superiority of half-frontal images is even more evident in this experiment, in which the reconstruction results are much closer to the ground truth landmarks.

4. CONCLUSION

We presented a novel 3D face reconstruction method that uses only a single face image with arbitrary pose as input. It combines a sparse 3D deformable model and a simple camera
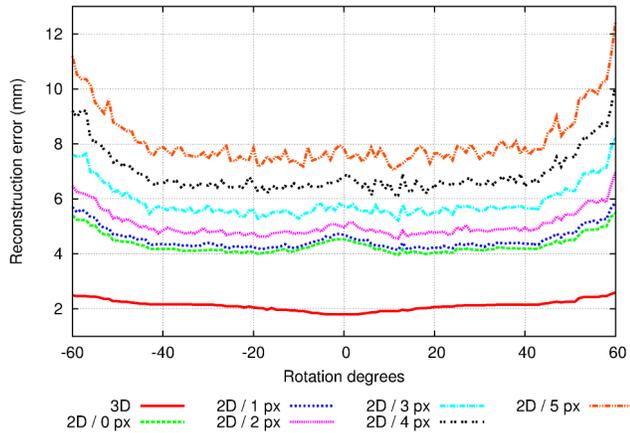
Fig. 6. Average reconstruction error of synthetic images with pose variation.
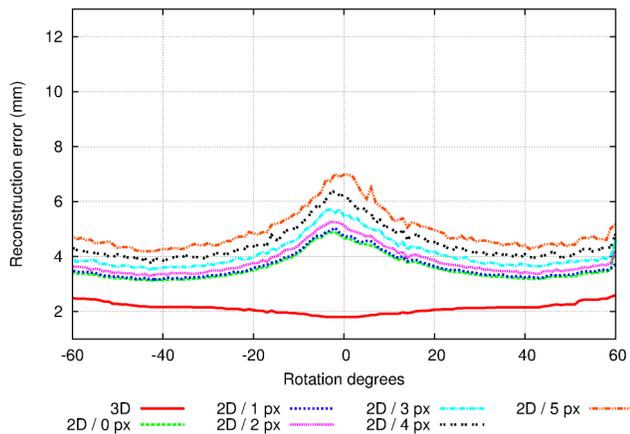
Fig. 7. Average reconstruction error of synthetic images with pose variation using the enhanced method.

model to estimate the 3D coordinates of 2D facial landmarks through iterative minimization of the reprojection error. We used the BU-3DFE database in our experiments to evaluate the reconstruction accuracy across pose variation and noise.

Corroborating previous neuropsychology works [16], our experiments have shown that half-frontal images present advantages over frontal images. Also, assuming that faces are symmetric improved the reconstruction accuracy and the robustness against large pose variations and noise. The average reconstruction error of half-frontal images ranged from 3.3 to $4.5mm$, depending on the noise, using only unknown subjects in the testing set.

As a future work, we intend to combine our reconstruction approach with TPS [2, 11] or to extend our sparse face model to a dense model [3] in order to obtain dense 3D models from a single face. We also intend to evaluate the performance of a face recognition system using only half-frontal face images and our 3D reconstruction method.

5. REFERENCES

[1] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *PAMI*, vol. 25, no. 9, pp. 1063–1074, 2003.

[2] J. Choi, G. Medioni, Y. Lin, L. Silva, O. Bellon, M. Pamplona Segundo, and T. C. Faltemier, "3D face reconstruction using a single or multiple views," in *ICPR*, 2010, pp. 3959–3962.

[3] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," *Journal of Pattern Recognition*, vol. 38, no. 6, pp. 787–798, 2005.

[4] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *PAMI*, vol. 33, no. 2, pp. 394–405, 2011.

[5] M. D. Levine and Y. Yu, "State-of-the-art of 3d facial reconstruction methods for face recognition based on a single 2d training image per person," *Pattern Recognition Letters*, vol. 30, no. 10, pp. 908–913, 2009.

[6] G. Medioni, J. Choi, C.-H. Kuo, and D. Fidaleo, "Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models," *SMC, Part A*, vol. 39, no. 1, pp. 12–24, 2009.

[7] 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," *PAMI*, vol. 33, no. 10, pp. 2115–2121, 2011.

[8] D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *PAMI*, vol. 23, no. 6, pp. 681–685, 2001.

[10] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *PAMI*, vol. 11, pp. 567–585, 1989.

[11] U. Park and A. K. Jain, "3d face reconstruction from stereo video," in *CRV 2006*, 2006, p. 41.

[12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[13] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[14] B. Bhanu and X. Zhou, "Face recognition from face profile using dynamic time warping," in *ICPR*, 2004, pp. 499–502.

[15] I. A. Kakadiaris, H. Abdelmunim, W. Yang, and T. Theoharis, "Profile-based face recognition," in *FG 2008*, 2008, pp. 1–8.

[16] G. Hole and V. Bourne, *Face Processing: Psychological, Neuropsychological, and Applied Perspectives*, Oxford, 1 edition, 2010.

[17] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *FG 2006*, 2006, pp. 211–216.

[18] I. T. Jolliffe, *Principal Component Analysis*, Springer, second edition, 2002.