Robust 3D Face Shape Reconstruction from Single Images via Two-Fold Coupled Structure Learning

Pengfei Dou
bensondou@gmail.com
Yuhang Wu
yuhang@cbl.uh.edu
Shishr K. Shah
sshah@central.uh.edu
Ioannis A. Kakadiaris
ioannisk@uh.edu

Computational Biomedicine Lab
Department of Computer Science
University of Houston
Houston, TX, USA

The problem of estimating the 3D shape of human faces from single images is of great interest and has attracted considerable research effort. Many approaches recently proposed to solve this problem could be considered extensions of Shape-from-Shading (SFS) methods, where a 3D shape is optimized to generate 2D renderings that match the input images [1, 5, 7]. Other methods in the literature propose to infer 3D face shape by fitting a set of feature points between the 2D image and the 3D model [3, 4, 6].

In this paper, we propose the Two-Fold Coupled Structure Learning (2FCSL) algorithm, which is capable of reconstructing 3D face models based on a sparse set of 2D landmarks that could be localized automatically by most of the recently proposed landmark detectors. By explicitly incorporating 3D-2D pose estimation and formulating the problem into a two-fold coupled structure learning problem, our method achieves better robustness to arbitrary pose variations and landmark localization noise.

Using a shape vector representation Y^i_{3D} of the dense 3D face, N 3D training faces are stacked together to construct the 3D dense landmark (3DDL) model $\gamma^d_{3D} = (Y^1_{3D}, \cdots, Y^N_{3D})$. Similarly, 3D sparse landmark (3DSL) model is represented by χ^s_{3D}=$(X^1_{3D}, \cdots, X^N_{3D})$, where X^i_{3D} is the vector representation of M 3D landmarks. Given a 2D image, a sparse set of landmarks X^I_{2D} is first detected with any off-the-shelf detector. Then, the 3D-2D projection matrix P is estimated using least squares minimization, such that $X^I_{2D} = P\bar{X}_{3D}$, where $\bar{X}_{3D}$ is the mean of 3DSLs in the training database. By projecting each 3DSL via P, the corresponding 2D sparse landmark (2DSL) model χ^s_{2D}=$(X^1_{2D}, \cdots, X^N_{2D})$, where X^i_{2D} is the vector representation of M 2D landmarks, is generated on-line.

By applying PCA to the 3DSL and the 2DSL models, we derive a compact representations of the corresponding shapes A_m and A_n, based on which a PLS regression P_{PLS} [2] is learned, $\hat{A}_m = A_n P_{PLS}$:

$$X^i_{3D} = \bar{X}_{3D} + \sum_{m=1}^{N-1} a^i_m U^s_{3D} \qquad A_m = [a^1_m, a^2_m, \cdots, a^{N-1}_m] \quad , \tag{1}$$

$$X^i_{2D} = \bar{X}_{2D} + \sum_{n=1}^{N-1} a^i_n U^s_{2D} \qquad A_n = [a^1_n, a^2_n, \cdots, a^{N-1}_n] \quad . \tag{2}$$

Following the same procedure, we compute the compact representation of X^I_{2D} by solving for $a^I_n = {U^s_{2D}}^{-1}(X^I_{2D} - \bar{X}_{2D})$. Then the a^I_m is recovered by $a^I_m = a^I_n P_{PLS}$ and the 3DSL is constructed through $X^R_{3D} = \bar{X}_{3D} + a^I_m U^s_{3D}$.

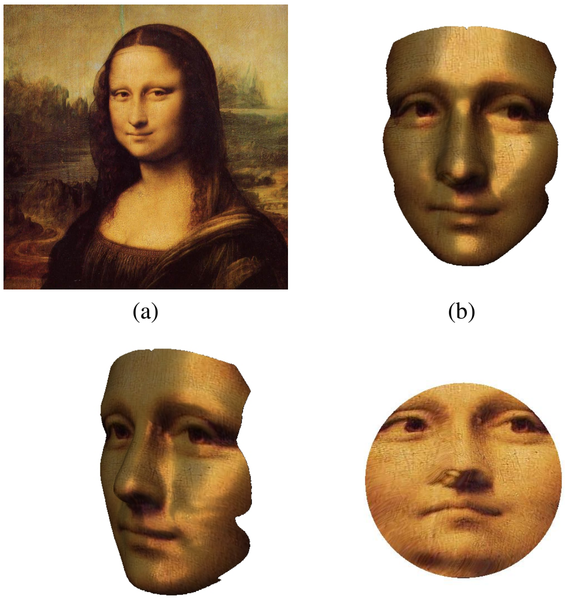
After we obtain the 3DSL X^R_{3D}, we aim to reconstruct the 3DDL Y^R_{3D}. In the training phase, the correlation between 3DSL and 3DDL is implicitly learned in a coupled manner.

$$\underset{\alpha, \Lambda^d_{3D}, \Lambda^s_{3D}}{\arg\min} \left\| \begin{bmatrix} \beta_0 \gamma^d_{3D} \\ \chi^s_{3D} \end{bmatrix} - \begin{bmatrix} \beta_0 \Lambda^d_{3D} \\ \Lambda^s_{3D} \end{bmatrix} \alpha \right\|^2_2 \quad s.t. \|\alpha\|_1 \leq \beta_1 \quad , \tag{3}$$

$$\underset{\alpha^*}{\arg\min} \left\| X^R_{3D} - \Lambda^s_{3D} \alpha^* \right\|^2_2 + \beta_2 \|\alpha^*\|_1 + \beta_3 \|\alpha^*\|_2 \quad . \tag{4}$$

By fitting X^R_{3D} to Λ^s_{3D}, the shared coefficient α^* could be recovered by solving Eq. 4. Then, the final Y^R_{3D} is reconstructed via Eq. 5:

$$Y^R_{3D} = \frac{\Lambda^d_{3D} \alpha^*}{\beta_0} \quad . \tag{5}$$



(c) (d)

Figure 1: (a) The input 2D image; (b) frontal view of the face shape reconstruction; (c) profile view of the face shape reconstruction; and (d) lifted UV texture.

In the paper, we conducted several experiments using both synthetic data and real 2D face images from two face datasets. Compared with [6], our method demonstrates higher reconstruction accuracy and better robustness to face pose variations and landmark localization noise. Fig. 1 depicts the reconstructed 3D face of Mona Lisa using the famous painting by Leonardo da Vinci and the lifted texture in a pre-registered UV space.

[1] R. Dovgard and R. Basri. *Statistical symmetric shape from shading for 3D structure recovery of faces*. Springer, 2004.

[2] P. Geladi and B.R. Kowalski. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185:1–17, 1986.

[3] T. Hassner. Viewing real-world faces in 3D. In *Proc. IEEE International Conference on Computer Vision*, Sydney, Australia, December 1-8 2013.

[4] I. Kemelmacher-Shlizerman and R. Basri. 3D face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2011.

[5] H. Rara, S. Elhabian, T. Starr, and A. Farag. Model-based shape recovery from single images of general and unknown lighting. In *Proc. IEEE International Conference on Image Processing*, pages 517–520, Cairo, November 2009.

[6] H.M. Rara, A.A. Farag, and T. Davis. Model-based 3D shape recovery from single images of unknown pose and illumination using a small number of feature points. In *Proc. International Joint Conference on Biometrics*, pages 1 –7, Washington, DC, October 2011.

[7] W.A.P. Smith and E.R. Hancock. Recovering facial shape using a statistical model of surface normal direction. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 28(12):1914–1930, December 2006.