# Disentangling Features in 3D Face Shapes for Joint Face Reconstruction and Recognition

Feng Liu[1], Ronghang Zhu[1], Dan Zeng[1], Qijun Zhao[1] and Xiaoming Liu[2]
[1]College of Computer Science, Sichuan University
[2]Department of Computer Science and Engineering, Michigan State University

In this supplementary material, we provide additional experimental results, including

- Face recognition results on IJB-A database;
- Phase-by-Phase Evaluation: CNN vs. 3DMM;
- Qualitative reconstruction results.

## 1. Recognition Results on IJB-A

The IJB-A database [4], including 5,396 images and 20,412 video frames of 500 subjects, has full pose variation and is more challenging than LFW [3]. We evaluate both face verification (1:1 comparison) and face identification (1:N search) performance of our proposed method with comparison to existing methods on the IJB-A database. The faces are firstly automatically detected by using the method in [7] and aligned by the method in [2]. If the automated methods fail, we manually crop the faces. The results are reported in Table 1.

When using only reconstructed shape features, our proposed method obtains the best face recognition accuracy in terms of true acceptance rate at false acceptance rate of $10\%$ (TAR-10%) and $1\%$ (TAR-1%), and rank-1 and rank-5 identification rate. Although it is outperformed by DR-GAN [5], a state-of-the-art texture-based face recognition method, the face recognition accuracy can be further improved after combining them by score-level summation fusion. These results, consistent with the results on the LFW and YTF [6] databases, prove the effectiveness of our proposed method in disentangling discriminative shape features that are complementary to texture features in face recognition as well as in surpassing the conventional 3D morphable model (3DMM) bases [1] in capturing facial detail.

Figure 1 shows some example genuine and imposter pairs in IJB-A, which are incorrectly recognized by DR-GAN [5], but correctly recognized by the fusion of DR-GAN and our proposed method. As can be seen, while extremely large head rotations may lead to the failure of existing texture-based face recognition methods, our proposed method explores complementary shape features to robustly recognize the off-angle faces with large rotations.

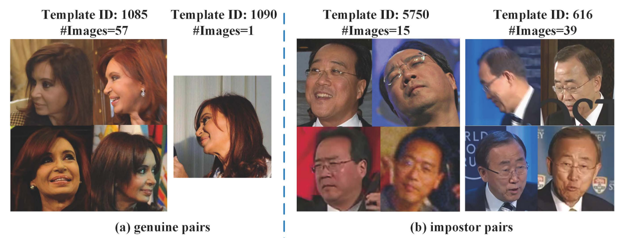


Figure 1. Example (a) genuine pairs and (b) imposter pairs in IJB-A, for which the state-of-the-art texture-based face recognition method (i.e., DR-GAN [5]) fails, whereas its fusion with our proposed method succeeds.

## 2. Phase-by-Phase Evaluation: CNN vs. 3DMM

Our proposed model is trained in three phases. Phases I and II replicate 3DMM for a proper initialization of our model, while Phase III makes our model beyond 3DMM by using joint supervisory of reconstruction and recognition (i.e., both reconstruction loss and identification loss). We compare the reconstruction and recognition results at different training phases. Table 2 gives the reconstruction results at Phases II and III, and summarizes the recognition results. It can be seen that reconstruction errors are further reduced after incorporating identification loss in Phase III. As for recognition, the accuracy is significantly improved from Phase II to Phase III. This reveals the limited discrimination power of 3DMM representations and the importance of CNN-based joint learning in expanding the representation and discrimination capacity of 3DMM-like bases.

Table 1. Face verification and identification performance on the IJB-A database.

| Method | Shape | Texture | TAR-10% | TAR-1% | Rank-1 | Rank-5 |
|---|---|---|---|---|---|---|
| | √ | × | $60.7 \pm 2.0$ | $30.6 \pm 3.2$ | $34.3 \pm 2.2$ | $55.1 \pm 2.1$ |
| 3DMM | × | √ | $71.1 \pm 1.8$ | $39.5 \pm 4.8$ | $49.8 \pm 2.5$ | $69.5 \pm 1.4$ |
| | √ | √ | $75.4 \pm 1.6$ | $46.6 \pm 5.1$ | $57.2 \pm 1.9$ | $74.4 \pm 1.3$ |
| 3DDFA | √ | × | $43.3 \pm 2.5$ | $12.5 \pm 1.9$ | $16.7 \pm 1.9$ | $38.3 \pm 2.7$ |
| | √ | × | $86.0 \pm 1.7$ | $55.9 \pm 5.5$ | $72.3 \pm 1.4$ | $88.0 \pm 1.4$ |
| 3DMM-CNN | × | √ | $83.5 \pm 2.2$ | $50.3 \pm 5.8$ | $70.9 \pm 1.5$ | $87.3 \pm 1.1$ |
| | √ | √ | $87.0 \pm 1.5$ | $60.0 \pm 5.6$ | $76.2 \pm 1.8$ | $89.7 \pm 1.0$ |
| DRGAN | × | √ | – | $75.5 \pm 2.8$ | $84.3 \pm 1.3$ | $93.2 \pm 0.8$ |
| Proposed | √ | × | $\mathbf{89.6 \pm 1.2}$ | $58.8 \pm 4.9$ | $75.7 \pm 1.9$ | $88.2 \pm 1.1$ |
| DRGAN+Proposed | √ | √ | – | $\mathbf{76.5 \pm 4.2}$ | $\mathbf{85.4 \pm 1.8}$ | $\mathbf{93.9 \pm 0.9}$ |

Table 2. Reconstruction and recognition accuracy on different test data sets when identity disentangling and identification loss are used or not used. Refer to the paper for test data set details.

| Training Phase | Identity Disentangling | Identification Loss | Reconstruction RMSE on | | | Recognition Accuracy on | |
|---|---|---|---|---|---|---|---|
| | | | MICC | BU3DFE (pose) | BU3DFE (exp.) | LFW | YTF |
| **–** | × | × | $2.51 \pm 0.57$ | $2.54 \pm 0.67$ | $2.62 \pm 0.73$ | **–** | **–** |
| **II** | √ | × | $2.23 \pm 0.48$ | $2.31 \pm 0.55$ | $2.45 \pm 0.62$ | $68.00 \pm 2.21$ | $69.19 \pm 1.91$ |
| **III** | √ | √ | $\mathbf{2.00 \pm 0.32}$ | $\mathbf{2.01 \pm 0.49}$ | $\mathbf{2.19 \pm 0.54}$ | $\mathbf{94.43 \pm 1.47}$ | $\mathbf{88.74 \pm 1.03}$ |

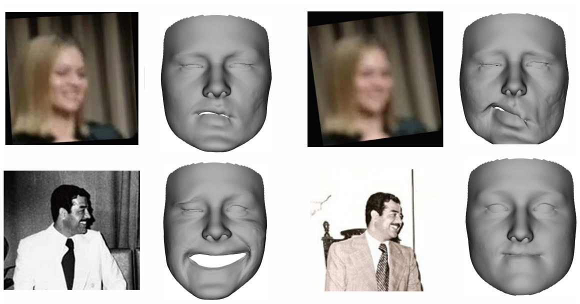

Figure 3. Failure cases of our proposed method due to blurry and very low resolution faces in the images/videos.

## 3. Qualitative Results

The 3D face reconstruction results of our proposed method on some images from the YTF and IJB-A databases are shown in Figure 2. One can obviously observe from these results that the reconstructed 3D faces do reveal the facial shape deformation (e.g., around the mouth), while the identity shapes successfully disentangle identity-sensitive from identity-irrelevant features. Figure 3 shows some images (video frames) for which our proposed method fails to generate plausible 3D face shapes. The blurry and very low resolution faces in these images/videos are the main reasons for the failure.

## References


[1] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *TPAMI*, 25(9):1063–1074, 2003.

[2] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). In *ICCV*, 2017.

[3] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[4] B. F. Klare, A. K. Jain, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and M. Burge. Pushing the frontiers of unconstrained face detection and recognition: IARPA janus benchmark A. In *CVPR*, pages 1931–1939, 2015.

[5] L. Tran, X. Yin, and X. Liu. Disentangled representation learning GAN for pose-invariant face recognition. In *CVPR, in press*, 2017.

[6] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR*, pages 529–534, 2011.

[7] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *SPL*, 23(10):1499–1503, 2016.

Figure 2. Reconstruction results by our proposed method on images from YTF (top) and IJB-A (bottom). The first row shows the input images, and the second and third rows show the reconstructed 3D shapes and *identity* shapes.