IDENTITY-INVARIANT FACIAL LANDMARK FRONTALIZATION FOR FACIAL EXPRESSION ANALYSIS

Vassilios Vonikakis

Advanced Digital Sciences Center (ADSC)
Singapore

*Stefan Winkler**

School of Computing
National University of Singapore (NUS)

ABSTRACT

We propose a frontalization technique for 2D facial landmarks, designed to aid in the analysis of facial expressions. It employs a new normalization strategy aiming to minimize identity variations, by displacing groups of facial landmarks to standardized locations. The technique operates directly on 2D landmark coordinates, does not require additional feature extraction and as such is computationally light. It achieves considerable improvement over a reference approach, justifying its use as an efficient preprocessing step for facial expression analysis based on geometric features.

1. INTRODUCTION

In the recent years, Facial Expression Analysis (FEA) has attracted strong attention, both in the research community [1], as well as in real-world systems and applications [2].

Faces however, are some of the most challenging objects to register. First, faces can exhibit a broad range of deformations which may significantly alter their shape and appearance, especially for high intensity expressions. Second, there is significant variability across individuals, resulting in considerable differences in appearance. This identity variability may further be accentuated by age, gender and race. Finally, perhaps the most important factor is the variability of faces across different viewpoints. As such, different combinations of yaw and pitch may result in dramatically different face appearance. Therefore, FEA systems operating in uncontrolled conditions have to at least address the head pose variability issue across different identities and expressions.

There are two main approaches to address this. The first option is to include extensive head pose variability during training, i.e. many training samples across different yaw-pitch combinations, allowing the model to directly learn pose-invariant face representations and exhibit some degree of robustness to different head poses. The straightforward approach to implement this is to use training datasets with many different viewpoints [3]. However, it is practically very difficult to obtain large datasets containing every possible facial action, across different demographic segments and under a diverse range of viewpoints. In such cases, synthetic 3D viewpoint augmentation on a 'frontal' dataset can introduce some degree of head pose variability [4, 5].

The second option is to include an explicit frontalization module, which can recover the frontal view of non-frontal facial images [6–9]. In this case, training may take place only on approximately frontal images, while during inference the frontalization module ensures that facial images are transformed to frontal before any analysis task.

Between these two approaches, frontalization modules offer more control in FEA, since they *decouple* head pose standardization and the expression analysis task. As such, they make it easier to build robust FEA systems, by utilizing only frontal or near-frontal datasets during training.

Frontalization can be classified into pixel- and landmark-based approaches. Pixel-based frontalization attempts to recover the frontal pixel appearance of the face. Usually this is achieved by fitting a 3D model on the non-frontal face, projecting pixels on the fitted model and rotating it back to frontal using texture warping [6–8, 10]. Recently, Generative Adversarial Networks (GANs) have been used successfully to reconstruct the frontal view of a non-frontal face [11–13]. Landmark-based frontalization on the other hand focuses on estimating only the location of facial landmarks as they would have looked from the front [9]; no appearance is estimated. As such, these approaches tend to be considerably less computationally expensive, since they do not require pixel rendering or generative networks. In recent years, directly estimating 3D facial landmarks from 2D images has become possible [14–17], simplifying the estimation of the frontal landmark view. However, these methods are still computationally expensive and not as widely used in FEA.

In many scenarios, light FEA approaches are required, e.g. for real-time execution in edge devices. In these cases, generative networks and expensive 3D reconstruction methods are usually not an option. Instead, FEA approaches solely based on geometric features that are derived directly from facial landmarks, represent a better choice. Even though no appearance features are used, it has been shown that they can

*The work was conducted while both authors were with the Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore.

achieve competitive results [3]. Therefore, using pixel-level frontalization (which is computationally expensive) followed by landmark detection may be overkill in these cases. Instead, landmark detection followed by landmark-based frontalization is a much faster approach.

In this paper we follow the latter approach. More specifically, we focus on the frontalization of faces in the context of FEA, aiming to discount two of the three major sources of variability of faces (identity and viewpoint), while preserving the third (deformations a.k.a. expressions). We also focus on a computationally light technique that allows for real-time frontalization of facial landmarks. To this end, we introduce a new fast facial landmark frontalization technique that compensates for viewpoint and identity variations in FEA tasks. The technique is data-driven and operates directly on the 2D coordinates of facial landmarks without the need for additional feature extraction. Once trained, it requires minimal overhead, since computation is reduced to a simple matrix multiplication. Experimental results demonstrate that it exhibits superior performance in comparison to other existing landmark-based frontalization approaches. As such, it may be used as preprocessing in landmark-based FEA systems, offering increased robustness to non-frontal headposes.

Our contributions can be summarized as follows:

- A new data-driven 2D facial landmark frontalization technique. It operates directly on the 2D facial landmark coordinates without requiring additional feature extraction. As such, it is very light, adding minimal computational overhead to FEA systems.

- A new landmark normalization strategy aiming to minimize identity variations, by displacing facial parts to standardized locations. This standardization strategy further improves the performance of the proposed frontalization approach.

2. 2D LANDMARK FRONTALIZATION

Given a facial image dataset of multiple viewpoints and expressions per subject, the main objective of the proposed method is to find a linear mapping from the 2D non-frontal landmarks of a given subject to the corresponding frontal landmarks of the *same* subject. At the same time, the frontal ground truth landmarks are standardized in terms of relative location of the major facial parts (i.e. eyes, nose, mouth), discounting any possible identity information.

Let ${}^k\mathbf{p}_i^j \in \mathbb{R}^{2N}$ be a vector containing the estimated x and y coordinates of N facial landmarks, from subject i with facial expression $j \in \{1, \cdots, J\}$ at viewpoint $k \in \{0, \cdots, K\}$. Let ${}^0\mathbf{p}_i^j$ denote the frontal viewpoint of the jth expression of subject i. Matrix $\mathbf{A}$ contains the transposed version of all landmark coordinate vectors ${}^k\mathbf{p}_i^j$, while $\mathbf{Y}$ is the 'ground truth' of their corresponding frontal viewpoint coordinates ${}^0\mathbf{p}_i^j$. With this, the required frontalization mapping $\mathbf{X}$ can be expressed through the following optimization:

$$\text{argmin}_{\mathbf{X}} \left\| \mathbf{Y} - \mathbf{A}\mathbf{X} \right\|_2 + \lambda \left\| \mathbf{X} \right\|_2, \tag{1}$$

which can be solved for $\hat{\mathbf{X}}$ in closed-form:

$$\hat{\mathbf{X}} = \left(\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I} \right)^{-1} \mathbf{A}^\top \mathbf{Y}, \tag{2}$$

where λ is the regularization term, and $\hat{\mathbf{X}} \in \mathbb{R}^{(2N+1)\times(2N)}$ represents the least-squares approximate solution, containing the frontalization weights for the mapping. It is important that matrices $\mathbf{A}$ and $\mathbf{Y}$ are filled in such a way that the landmark coordinates of every given non-frontal expression of a subject are mapped to the landmark coordinates of the corresponding frontal expression of the *same* subject. In the following example, single horizontal lines represent the transition to another expression of the same subject, while double horizontal lines represent the transition to a different subject.

$$\mathbf{A} = \begin{bmatrix} 1 & {}^1\mathbf{p}_1^{1\top} \\ \vdots & \vdots \\ 1 & {}^K\mathbf{p}_1^{1\top} \\ \hline \vdots & \vdots \\ \hline 1 & {}^1\mathbf{p}_1^{J\top} \\ \vdots & \vdots \\ 1 & {}^K\mathbf{p}_1^{J\top} \\ \hline\hline 1 & {}^1\mathbf{p}_2^{1\top} \\ \vdots & \vdots \\ 1 & {}^K\mathbf{p}_2^{1\top} \\ \hline \vdots & \vdots \\ \hline 1 & {}^1\mathbf{p}_2^{J\top} \\ \vdots & \vdots \\ 1 & {}^K\mathbf{p}_2^{J\top} \\ \hline\hline \vdots & \vdots \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} {}^0\mathbf{p}_1^{1\top} \\ \vdots \\ {}^0\mathbf{p}_1^{1\top} \\ \hline \vdots \\ \hline {}^0\mathbf{p}_1^{J\top} \\ \vdots \\ {}^0\mathbf{p}_1^{J\top} \\ \hline\hline {}^0\mathbf{p}_2^{1\top} \\ \vdots \\ {}^0\mathbf{p}_2^{1\top} \\ \hline \vdots \\ \hline {}^0\mathbf{p}_2^{J\top} \\ \vdots \\ {}^0\mathbf{p}_2^{J\top} \\ \hline\hline \vdots \end{bmatrix} \tag{3}$$

In order to train the frontalization module, 5 different datasets are employed, all containing portrait images of different people with varying facial expressions under multiple viewpoints. Table 1 depicts the datasets used for training the frontalization module, along with their characteristics. All available images from the 5 datasets were scanned with OpenCV's frontal and profile face detectors. All valid detected faces were analyzed using Supervised Descent Method (SDM) [18] in order to estimate the 2D facial landmarks. The total number of valid detected faces was 86844, which were doubled to 173688 after mirroring. Each set of raw coordinates was separately normalized for scaling, translation and rotation (see Section 2.1), before filling the $\mathbf{A}$ and $\mathbf{Y}$ matrices and estimating the frontalization weights $\hat{\mathbf{X}}$.

The computational overhead introduced by the proposed frontalization approach at runtime is minimal, since for N

Table 1. Datasets used for training the frontalization module.

Dataset	Pitch	Yaw
Radboud [19]	–	0°, ±45°, ±90°
Karolinska [20]	–	0°, ±45°, ±90°
CAS-PEAL [21]	±15°	0°,±22°,±45°, ±67°,±90°
PIE [22]	±15°	0°,±22°,±45°, ±67°,±90°
Multi-PIE [23]	−30°	0°,±15°,±30°,±45°, ±60°,±75°,±90°

facial landmarks, this equates to a single vector-matrix multiplication between the coordinate vector $\mathbf{p} \in \mathbb{R}^{2N+1}$ ($2N$ landmark coordinates plus the interception) and the frontalization matrix $\hat{\mathbf{X}} \in \mathbb{R}^{(2N+1)\times(2N)}$.

2.1. Normalization of landmarks

Let $\mathbf{P}_i^N \in \mathbb{R}^{N\times 2}$ be a matrix containing the x and y coordinates of N facial landmarks, of facial image i. In order to properly learn the frontalization matrix $\hat{\mathbf{X}}$, all coordinates in matrices $\mathbf{A}$ and $\mathbf{Y}$ should be appropriately standardized for translation, scaling and rotation. For this, we employ a non-isotropic version of the Procrustes transformation.

Let $n \subset N$ represent a subset of the *less-deformable* facial landmarks; landmarks that change minimally with different expressions. These include facial points on the eyes, the nose, and the middle of the upper lip. Eyebrows and the majority of the mouth are excluded, since their coordinates may change significantly with different expressions, thus, introducing additional variations that may affect scale. In our case, $N = 49$ and $n = 13$. Let $\mathbf{P}_i^n \in \mathbb{R}^{n\times 2}$ represent the 'reduced' coordinate matrix of the less-deformable n points. Then, the standardized coordinate vector $\hat{\mathbf{P}}_i^N$ is estimated as follows:

$$\hat{\mathbf{P}}_i^N = \left(\frac{\left\|\mathbf{P}_i^n - {}^n\overline{\mathbf{P}}_i^n\right\|_F}{\sqrt{n}}\right) \cdot \left(\mathbf{P}_i^N - {}^N\overline{\mathbf{P}}_i^n\right) \cdot \mathbf{R}_i \quad (4)$$

where ${}^x\overline{\mathbf{P}}_i^n \in \mathbb{R}^{h\times 2}$ denotes a matrix containing the mean values of the columns of $\mathbf{P}_i^n$, replicated for x rows. $\|\cdot\|_F$ represents the Frobenius norm, while $\mathbf{R}_i \in \mathbb{R}^{2\times 2}$ is the following rotation matrix:

$$\mathbf{R}_i = \begin{bmatrix}\cos(a) & -\sin(a)\\ \sin(a) & \cos(a)\end{bmatrix}, a = \text{atan}\left(\frac{x_{eyeR} - x_{eyeL}}{y_{eyeR} - y_{eyeL}}\right)$$

with x_{eye}, y_{eye} representing the coordinates of the centers of the left or right eyes of image i.

Equation (4) has 3 distinct terms, standardizing the scaling, translation and rotation, respectively. The scaling factor essentially is the average Euclidean distance of the less-deformable landmarks to their corresponding mean. The fact that only the mean values of the less-deformable points n are used for the calculation of the scaling and translation factors adds robustness to different facial expressions. Conversely, if all the N landmarks were used, then various facial expressions would affect the scale and the translation of the standardized landmark coordinates. This approach is also more robust to yaw changes, compared to the widely used intra-ocular (eye-to-eye) distance, since in non-frontal faces, the eye-to-eye distance becomes smaller, and thus, normalizing by it affects the scaling term.

2.2. Identity standardization

Although the above standardization approach eliminates variabilities related to scale, translation and rotation, a person's identity may still introduce unnecessary variations that can affect the frontalization performance. This is due to the fact that every person has different relative positions of landmark groups, e.g. distance between the eyes, distance of the upper mouth to the bottom of the nose, upper starting point of the nose etc. During our experiments we found out that eliminating this type of identity variability further improves the performance of the frontalization module.

To this end, we introduced additional translation terms *only* to the frontal ground truth coordinates of matrix $\mathbf{Y}$. These translation terms displace whole groups of facial landmarks, such as eyes, mouth and nose, and re-position them in the corresponding group locations of a *template* face. By doing this, we force the optimization to learn a more *identity-invariant* frontalization transformation, from non-frontal landmarks towards a standardized frontal face, discounting individual face variabilities.

Let $\{\mathcal{E}_L, \mathcal{E}_R, \mathcal{M}, \mathcal{N}\} \subset N$ represent subsets of facial landmarks, corresponding to the left and right eye, the mouth and the nose, respectively. Let also $\mathbf{P}_i^{\mathcal{E}_L}$, $\mathbf{P}_i^{\mathcal{E}_R}$, $\mathbf{P}_i^{\mathcal{M}}$, $\mathbf{P}_i^{\mathcal{N}}$ represent the landmark coordinates of these subsets for facial image i, while $\mathbf{P}_T^{\mathcal{E}_L}$, $\mathbf{P}_T^{\mathcal{E}_R}$, $\mathbf{P}_T^{\mathcal{M}}$, $\mathbf{P}_T^{\mathcal{N}}$, the landmark coordinates of these subsets for the template face. Then the identity-standardized frontal coordinates $\tilde{\mathbf{P}}_i^x$ of matrix $\mathbf{Y}$ are given by the following equations:

$$\tilde{\mathbf{P}}_i^x = \hat{\mathbf{P}}_i^x + \Delta\hat{\mathbf{P}}_i^x \quad (5)$$
$$\text{where } \Delta\hat{\mathbf{P}}_i^x = \bar{\hat{\mathbf{P}}}_T^x - \bar{\hat{\mathbf{P}}}_i^x$$

with $x \in \{\mathcal{E}_L, \mathcal{E}_R, \mathcal{M}, \mathcal{N}\}$ and $\bar{\hat{\mathbf{P}}}^x$ representing an *anchor* point in the landmarks of subset x. For subsets $\mathcal{E}_L, \mathcal{E}_R$ and $\mathcal{N}$, the anchor point used was the mean of their standardized coordinates $\hat{\mathbf{P}}_i^x$. For $\mathcal{M}$, the upper middle landmark was used as anchor point, since different mouth deformations can affect its coordinate mean considerably, introducing fluctuations. Matrix $\Delta\hat{\mathbf{P}}_i^x$ adds a displacement in the standardized coordinates, such that the different landmark subsets will always be located in the same location as in the template face T. In our experiments, the mean frontal face of the training datasets was used as template face. It is important to highlight

that the displacement term does not affect the shape of the landmark subsets, but only their *relative* position within the face. For example, a smiling mouth will maintain its shape, and only its position within the face will change. Also, this type of identity standardization takes place only once, during training of the frontalization module. In runtime, all facial landmarks are standardized only for scale, rotation and translation, with equation (4), while the frontalization module will estimate their frontal view, suppressing the individual identity differences and maintaining the expression.

3. EXPERIMENTAL RESULTS

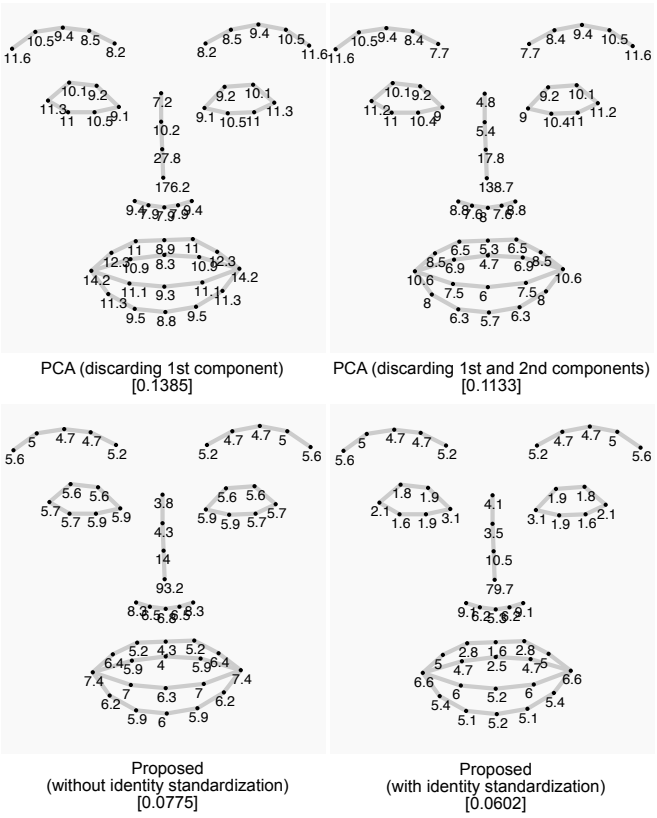

Fig. 1. Total average errors [in brackets] and per landmark average errors of different frontalization approaches (not showing $\times 10^{-2}$).

In order to quantitatively evaluate the performance of the proposed frontalization approach, a frontalization module was trained using 4 out of the 5 datasets of Table 1 (Radboud, Karolinska, CAS-PEAL, PIE) and tested on the holdout dataset (Multi-PIE). The normalized least square residual $\left\| f({}^k\mathbf{p}_i^j) - {}^0\mathbf{p}_i^j \right\| / \left\| {}^0\mathbf{p}_i^j \right\|$ was used to measure frontalization performance, which is the same metric used in face alignment [18]. $f({}^k\mathbf{p}_i^j)$ represents the frontalized view of the non-frontal landmarks ${}^k\mathbf{p}_i^j$, while ${}^0\mathbf{p}_i^j$ its actual ground truth frontal landmarks, provided by the test set. Fig. 1 depicts the frontalization errors of two versions of the proposed method (with and without identity standardization), as well as that of a PCA-based approach [9]. According to this approach, the first principal components, with the largest capacity to explain variance, are associated with yaw and pitch. Consequently, if discarded during reconstruction, yaw and pitch are discounted and thus, the landmarks are frontalized. Errors are depicted both as average per landmark (across all images), as well as total average (across all images and landmarks). It is evident that the proposed approach exhibits considerably smaller errors compared to the PCA-based approach. Additionally, identity standardization can further reduce the error across all landmarks, resulting in a more stable frontalized face.

Fig. 2 depicts example outputs of the final frontalization approach (trained with all 5 datasets of Table 1) on *unseen* images captured by a web camera. Based on our qualitative tests, the frontalization module is quite robust to yaw ranges between $\pm 45^\circ$ and pitch ranges between $\pm 15^\circ$. After that performance drops, but still remains good for yaw ranges in $[\pm 45^\circ, \pm 60^\circ]$ and pitch ranges in $[\pm 15^\circ, \pm 30^\circ]$. It was not possible to test the performance above $\pm 60^\circ$ for yaw and $\pm 30^\circ$ for pitch, since this is the operating range of the landmark detector and thus, faces with more extreme head poses would either yield unreliable facial landmarks or would not be detected by the face detector.

We also observe that there is an imbalance in performance between yaw and pitch. This could be due to 2 possible reasons. First, the datasets that we used had considerably fewer training examples for pitch variations compared to yaw. Second, yaw frontalization is inherently easier compared to pitch, since the natural left-right symmetry of the face helps with it.

4. CONCLUSION

We presented a technique for frontalizing 2D facial landmarks, which is designed for facial expressions analysis. Our approach employs a new point normalization strategy that aims to minimize identity variations and shifts different facial landmarks to standard locations. The technique operates directly on 2D landmark coordinates and does not require additional feature extraction. As such, it adds minimal computational overhead making it suitable for real-time systems. Benchmarking with several face datasets shows that it outperforms a PCA-based reference approach by a substantial margin, approximately halving average errors.

5. REFERENCES

[1] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pat-*

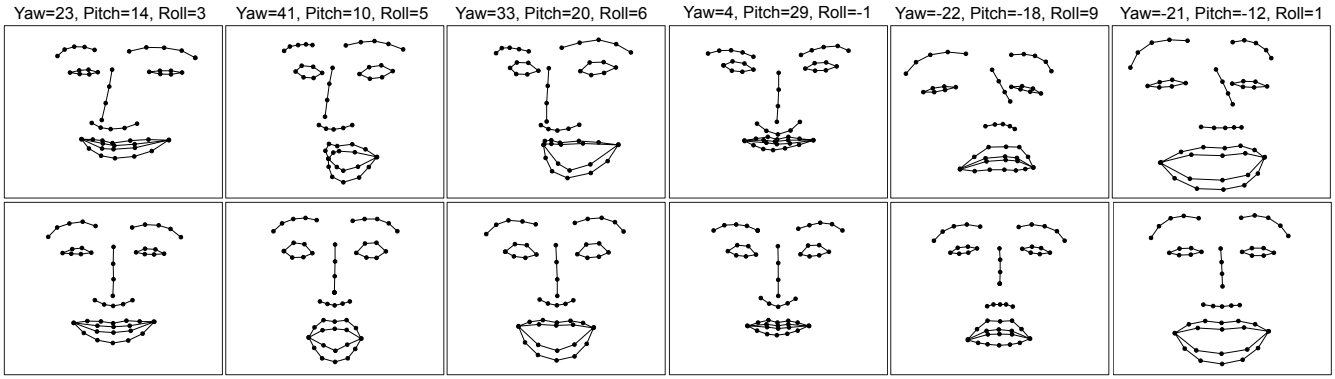

Fig. 2. Example outputs of the frontalization module on *unseen* non-frontal images of various head poses and expressions. Top: Original landmarks and head pose estimated by SDM. Bottom: Output of the proposed frontalization module.

tern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1113–1133, June 2015.

[2] V. Vonikakis and S. Winkler, "Emotion-based sequence of family photos," in *Proc. 20th ACM International Conference on Multimedia*. ACM, 2012, p. 1371–1372.

[3] V. Vonikakis, Y. Yazici, V. D. Nguyen, and S. Winkler, "Group happiness assessment using geometric features and dataset balancing," in *Proc. 18th ACM International Conference on Multimodal Interaction (ICMI)*. ACM, 2016, pp. 479–486.

[4] I. Masi, A. T. Tran, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in *Proc. European Conf. Computer Vision (ECCV)*, 2016, pp. 579–596.

[5] D. E. Crispell, O. Biris, N. Crosswhite, J. Byrne, and J. L. Mundy, "Dataset augmentation for pose and lighting invariant face recognition," *CoRR*, vol. abs/1704.04326, 2017.

[6] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4295–4304.

[7] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo, "Effective 3D based frontalization for unconstrained face recognition," in *Proc. 23rd International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 1047–1052.

[8] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 1701–1708.

[9] M. F. Valstar, E. Sánchez-Lozano, J. F. Cohn, L. A. Jeni, J. M. Girard, Z. Zhang, L. Yin, and M. Pantic, "FERA 2017 – addressing head pose in the third facial expression recognition and analysis challenge," in *Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2017, pp. 839–847.

[10] W. Deng, J. Hu, Z. Wu, and J. Guo, "Lighting-aware face frontalization for unconstrained face recognition," *Pattern Recognition*, vol. 68, pp. 260–271, 2017.

[11] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Towards large-pose face frontalization in the wild," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 4010–4019.

[12] Y. Qian, W. Deng, and J. Hu, "Unsupervised face normalization with extreme pose and expression in the wild," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 9843–9850.

[13] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1283–1292.

[14] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3D face alignment from 2D video for real-time use," *Image and Vision Computing*, vol. 58, pp. 13–24, 2017.

[15] S. Tulyakov, L. A. Jeni, J. F. Cohn, and N. Sebe, "Viewpoint-consistent 3D face alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 9, pp. 2250–2264, Sept 2018.

[16] 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 *Proc. IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 1021–1030.

[17] R. Zhao, Y. Wang, and A. M. Martinez, "A simple, fast and highly-accurate algorithm to recover 3D shape from 2D landmarks on a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3059–3066, Dec 2018.

[18] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, pp. 532–539.

[19] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud faces database," *Cognition and Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[20] E. Goeleven, R. D. Raedt, L. Leyman, and B. Verschuere, "The Karolinska directed emotional faces: A validation study," *Cognition and Emotion*, vol. 22, no. 6, pp. 1094–1118, 2008.

[21] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, no. 1, pp. 149–161, Jan 2008.

[22] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, Dec 2003.

[23] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.