Face Frontalization for Cross-Pose Facial Expression Recognition

Deniz Engin*, Christophe Ecabert†, Hazım Kemal Ekenel*, Jean-Philippe Thiran†
* Istanbul Technical University (ITU), Turkey
† École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
deniz.engin@itu.edu.tr, christophe.ecabert@epfl.ch, ekenel@itu.edu.tr, jean-philippe.thiran@epfl.ch

Abstract—**In this paper, we have explored the effect of pose normalization for cross-pose facial expression recognition. We have first presented an expression preserving face frontalization method. After face frontalization step, for facial expression representation and classification, we have employed both a traditional approach, by using hand-crafted features, namely local binary patterns, in combination with support vector machine classification and a relatively more recent approach based on convolutional neural networks. To evaluate the impact of face frontalization on facial expression recognition performance, we have conducted cross-pose, subject-independent expression recognition experiments using the BU3DFE database. Experimental results show that pose normalization improves the performance for cross-pose facial expression recognition. Especially, when local binary patterns in combination with support vector machine classifier is used, since this facial expression representation and classification does not handle pose variations, the obtained performance increase is significant. Convolutional neural networks-based approach is found to be more successful handling pose variations, when it is fine-tuned on a dataset that contains face images with varying pose angles. Its performance is further enhanced by benefiting from face frontalization.**

Index Terms—**Expression preserving face frontalization, cross-pose facial expression recognition, convolutional neural networks**

I. Introduction

Facial expression plays a significant role in human-human and human-machine interactions. Many real world applications require automatic facial expression recognition (FER) under unconstrained environments. Hence, facial images appearing in various poses should be handled by the developed FER systems. One way of addressing this problem is to perform face frontalization to eliminate the variations in facial appearance due to different views. However, it is critical to preserve facial expression, while performing pose normalization.

Pose normalization has been used for cross-pose face recognition [1]–[9]. On the other hand, these methods have not been applied for facial expression recognition, since they were mainly focusing on the identification aspects and were not specifically developed for preserving facial expressions. Existing pose normalization studies on face recognition can be based on, for example, 2D methods [1]–[5] and 3D methods [6]–[8]. In [9], homography-based normalization (HPN) has been proposed utilizing both 2D and 3D methods to be able to extract features according to each pose, also 3D facial landmarks were mapped to each 2D face image.

Previous cross-pose facial expression recognition studies have mainly focused on learning transformations in the feature space due to varying pose. Transductive transfer linear discriminant analysis (TTLDA) method has been proposed to transfer the label information from source facial pose to unlabeled auxiliary target facial pose in [10]. In [11], dictionary learning based framework has been introduced to learn a cross-modality and pose-invariant dictionary that contains 3D shape and 2D texture information. Partial Least Squares (PLS) method has been used to learn relations between pose pairs, which belong to the same individual and same facial expression from different pose angles in [12]. In [13], head pose normalization based on given 2D locations has been introduced, in addition, proposed Gaussian Process Regression model has been applied to learn relationships between different pose angles.

In recent years, besides hand-crafted features such as local binary pattern (LBP), scale invariant feature transform (SIFT), histogram of gradients (HoG), and Gabor filters, deep learning-based methods, commonly convolutional neural networks (CNNs), have been widely used in computer vision, such as for object detection, object recognition, face and facial expression recognition. AlexNet [14], VGG [15], GoogLeNet [16], and ResNet [17] are among the well-known and successful architectures for image classification. Therefore, in this study, after the face frontalization step, for facial expression representation and classification, we have employed both a traditional approach, by using hand-crafted features, local binary patterns, in combination with support vector machine classification and a convolutional neural networks-based approach.

In this work, our main aim is to improve cross-pose facial expression recognition by employing facial expression preserving face frontalization. To evaluate the impact of face frontalization on facial expression recognition performance, we have conducted cross-pose FER experiments using the BU3DFE database [18]. The database consists of 3D models from 100 individuals. Six basic expressions have four level of intensity, thus six basic expressions are classified at different intensity levels. Firstly, we apply pose normalization on non-frontal face images, and train a model by using LBP features with support vector machine (SVM) classifiers, and CNN for facial expression classification. Experimental results show that face frontalization increases cross-pose FER accuracy, especially, when LBP features in combination with SVM

ISBN 978-90-827970-1-5 © EURASIP 2018

classifiers are used. Since this facial expression representation and classification does not handle pose variations, the obtained performance improvement is very significant with 17% absolute increase in average accuracy. CNN-based approach is found to be more successful handling pose variations, when it is fine-tuned on a dataset that contains face images with varying pose angles. Its performance is further improved by benefiting from the face frontalization. The obtained absolute increase in average accuracy is 1.2%. The contributions and main outcomes of this paper can be listed as below:

- We have investigated the effect of pose normalization for cross-pose FER and have presented an expression preserving face frontalization method.
- We have assessed the impact of face frontalization both on a traditional FER approach, by using LBP features in combination with SVM classification, and on a deep learning-based approach.
- We have found that face frontalization improves cross-pose FER performance significantly, especially, when the face representation is not robust against pose variations.
- CNN-based approach has been found to be more successful in handling pose variations, when it is fine-tuned on a dataset that contains face images with varying pose angles.

The rest of the paper is organized as follows: In Section II, the proposed method is described in detail. Experimental results are presented and discussed in Section III. Finally, the conclusions are given in Section IV.

II. METHODS

In this section, the face frontalization process and facial expression recognition methods are presented.

A. Pose Normalization

This section describes all the steps involved in the generation of a virtual frontal view from multiple cameras. The process reconstructs the shape at first, then blends the textures coming from different views into a common texture map. The final image can be rendered using standard computer graphic pipeline.

1) Notation: Following notation is used in this paper: Italic capital letter $\mathscr{T}$ denotes tensor, bold capital letter $\mathbf{M}$ denotes matrix, bold lower-case letter $\mathbf{v}$ denotes vector, and italic lower-case letter s stands for scalar value. Operation on tensor such as mode-n product is denoted by $\mathscr{T} \times_n \mathbf{M}$ or $\mathscr{T} \times_n \mathbf{v}$.

2) 3D Face Model: Most of the *morphable models* used nowadays are linear and usually capture a single attribute [19]. Building one that can address different types of variations, such as identity and expression, increases the complexity and difficulty. This process is simplified by assuming that the identity and expression can be decoupled and each attribute can be modeled on their own. Such decoupling can be achieved by using the multilinear model as shown in [20].

The *bilinear face model* used in this work has been trained using the FaceWarehouse database introduced in [21]. An arbitrary face $\mathbf{f}$ can be generated as:

$$\mathbf{f} = \mathscr{C}_r \times_2 \mathbf{w}_{id}^\top \times_3 \mathbf{w}_{exp}^\top \tag{1}$$

where $\mathscr{C}_r$ is the tensor model, $\mathbf{w}_{id}^\top$ and $\mathbf{w}_{exp}^\top$ are the weights for identity and expression, respectively.

3) Fitting Method: The fitting method [22], which is based on detected facial landmarks, relies on two assumptions: 1) There is a direct correspondence between vertices (*i.e. 3D Points*) and facial landmarks located in the image space; 2) The 3D surface can be recovered using only a small set of control points (*i.e. sparse measurements*) [23]. Given a sparse measurement vector $\mathbf{r}_i \in \mathbb{R}^{2f}$ composed of f facial landmarks for the i^{th} image, equation (2) defines the relationship between the reconstructed 3D surface and the facial landmarks, where $\mathbf{P}_i$ represents the sparse vertices selection going from N to f vertices ($\mathbf{P}_i : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3f}$) and $\mathbf{L}_i$ models the projection operator [22].

$$\mathbf{r}_i = \mathbf{L}_i \mathbf{P}_i \mathbf{f} = \mathbf{L}_i \mathbf{P}_i \left(\mathscr{C}_r \times_2 \mathbf{w}_{id}^\top \times_3 \mathbf{w}_{exp}^\top \right). \tag{2}$$

The identity and expression weights can be estimated by minimizing equation (3), the error between the projected control points and the landmarks in each view.

$$E_{data} = \frac{1}{N} \sum_{i=0}^{K} \left\| \mathbf{L}_i \mathbf{P}_i \left(\mathscr{C}_r \times_2 \mathbf{w}_{id}^\top \times_3 \mathbf{w}_{exp}^\top \right) - \mathbf{r}_i \right\|^2 \tag{3}$$

Two additional terms E_{id}, E_{exp} are added to penalize the magnitude of the deformation.

$$E_{id} = \left\| \mathbf{w}_{id}^\top \right\|^2 \qquad E_{exp} = \left\| \mathbf{w}_{exp}^\top \right\|^2 \tag{4}$$

Finally, the objective function defined in equation (5) can be solved using coordinate-descent approach to estimate identity and expression parameters [20].

$$E = E_{data} + \eta_{id} E_{id} + \eta_{exp} E_{exp} \tag{5}$$

In our experiments, setting η_{id} and η_{exp} to 0.5 and 0.1, respectively showed satisfactory results.

4) Texture Generation: Color information from each view can be fused in the texture map using a similar approach as in [24]. Using the reconstructed surface, each view v_i is warped into a common texture space $\mathbf{I}_i^{warp}$. For each pixel in the texture map $\mathbf{I}_p$, the color is computed by weighting the contribution of each warped pixels. For each vertex $\mathbf{v}_k$, a weight $w_i(\mathbf{v}_k)$ proportional to its visibility from the view v_i is defined:

$$w_i(\mathbf{v}_k) = e^{-\omega_n (1 - \mathbf{n}_k \mathbf{n}_i^{cam})} \tag{6}$$

where $\mathbf{n}_k$ is the normal for $\mathbf{v}_k$ and $\mathbf{n}_i^{cam}$ represents the $\mathbf{z}$ direction of the i^{th} camera. The weight $w_i(\mathbf{v}_k)$ is propagated to pixel by interpolating them with radial basis function, using the following equation:

$$w_{i,p} = \sum_k e^{\omega_u ||\mathbf{u}_p - \mathbf{u}_{i,k}||^2} w_i(\mathbf{v}_k) \tag{7}$$

Fig. 1. Sample outputs of face frontalization process, from left to right, for 0, 15, 30, 45, 60 degrees, respectively. First two rows show different views generated for a given subject, whereas the last row shows the frontalized face images. Up to medium angles (*i.e.* 45), the presented method is able to generate a frontal face image successfully. However with larger pose angles, discontinuities start to show in the middle of the face, due to self occlusion and erroneous landmarks localization. For the FER experiments, we only used the pose angles up to 45 degrees.

where $\mathbf{u}_p$ is the coordinate of the pixel p, $\mathbf{u}_{i,k}$ is the position of $\mathbf{v}_k$ in texture space and ω_u controls the support region of $\mathbf{v}_k$. Each pixel's weight is then normalized such that $\sum_i w_{i,p} = 1$. The final pixel value in the texture map is given by:

$$\mathbf{I}_p = \sum_i w_{i,p} \mathbf{I}_{i,p}^{warp} \tag{8}$$

In our experiments, ω_n and ω_u have been empirically set to 50 and $1e^{-3}$, respectively. In Fig. 1, the first two rows show different views generated for a given subject. Starting with no pose, it is gradually increased by steps of 15 degrees up to 60 degrees to the right and left, respectively. Due to self occlusions, we limit the pose angles up to 60 degrees and for the FER experiments, use the pose angles up to 45 degrees. The last row shows the pose normalized face images. Up to medium angles, *i.e.* 45 degrees, the solution presented in this section is able to generate a frontal image successfully. However with larger pose, discontinuities start to show in the middle of the face, due to self occlusion and wrong landmark localization.

B. Facial Expression Recognition

We have performed facial expression recognition both with hand-crafted features, LBP, combined with support vector machine classifiers, and convolutional neural networks. In this subsection, we give an overview of these approaches. Parameters are provided in *III. B. Experimental Setup* subsection.

1) Local Binary Patterns: We have used spatially enhanced LBP histogram [25] for face representation. In this approach, images are divided into regions and uniform LBP histogram is calculated for each region. The overall feature vector is obtained by concatenating histograms from each region.

2) Convolutional Neural Networks: In the CNN-based approach, we have employed the VGG-Face architecture [26]. VGG-Face model has 13 convolutional layers followed by three fully connected layers.

Fig. 2. Sample images from the FER 2013 database [27].

III. EXPERIMENTAL RESULTS

In this section, we first present the databases used in this work and the experimental setup. Then, we convey the experimental results and discuss them in detail.

A. Databases

1) BU3DFE Database: The Binghamton University 3D Facial Expression Database (BU3DFE) [18] is used in this study. This database contains 3D models from 100 subjects, 56 females and 44 males. 83 landmark points are provided for each model. The subjects show a neutral face as well as six basic facial expressions —anger, disgust, fear, happiness, sadness, surprise— at four different intensity levels.

2) FER2013 Database: The Facial Expression Recognition 2013 [27] database was released for the ICML 2013 Challenge on Representation Learning. It was collected by using Google image search, and contains a large amount of appearance variations, mainly due to illumination and pose. It contains 35,887 images with six basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. Samples from this database can be seen in Fig. 2.

B. Experimental Setup

We designed a cross-pose, subject-independent experimental setup where the 100 subjects are divided into five folds according to subject IDs. Using these subject-independent subsets, we performed 5-fold cross validation. Training data includes images of 80 subjects, while test data includes images of 20 subjects. Training data has only zero degree (frontal) face images, whereas test data contains face images with different pose angles varying from -45 to 45 degrees in 15 degrees steps. Different intensity levels of expressions are trained and tested, separately.

We have chosen $LBP_{(8,2)}^{u2}$ operator for LBP-based feature extraction process. Each image is divided into $7 \times 7 = 49$ regions. In uniform LBP, each region is represented by 59 dimensional features. For the whole image this results in $59 \times 49 = 2891$ dimensional feature vector. After feature extraction step, principal component analysis (PCA) is applied, and 99% of the variance is retained. This way, feature dimensionality is reduced to 400. Finally SVMs with radial basis function (RBF) kernel are trained in one-vs-all manner. LIBSVM library is used for implementation [28].

For CNN-based approach, we have used the pre-trained VGG-Face model [26] and fine-tuned it on the FER 2013

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE I
RECOGNITION ACCURACIES ON NORMALIZED IMAGES AT DIFFERENT INTENSITIES BY USING LBP.

Intensity	15	30	45	Avg.
1	57.43	53.92	48.81	53.39
2	67.33	65.50	54.68	62.50
3	71.50	65.76	53.09	63.45
4	76.33	73.16	59.41	69.63
Avg.	68.15	64.59	54.00	**62.24**

TABLE III
RECOGNITION ACCURACIES ON NORMALIZED IMAGES AT DIFFERENT INTENSITIES BY USING CNN.

Intensity	15	30	45	Avg.
1	62.83	59.33	55.67	59.28
2	69.33	64.17	59.50	64.33
3	75.17	73.17	67.67	72.00
4	80.17	76.17	68.00	74.78
Avg.	71.87	68.21	62.71	**67.60**

TABLE II
RECOGNITION ACCURACIES ON NON-NORMALIZED IMAGES AT DIFFERENT INTENSITIES BY USING LBP.

Intensity	45l	30l	15l	15r	30r	45r	Avg.
1	27.83	34.67	45.17	49.00	40.50	28.67	37.64
2	23.83	41.17	57.00	59.33	48.67	35.83	44.31
3	33.50	52.16	63.83	65.50	49.50	29.83	49.05
4	32.83	57.83	68.00	63.50	44.83	30.50	49.58
Avg.	29.50	46.46	58.50	59.33	45.88	31.21	**45.15**

TABLE IV
RECOGNITION ACCURACIES ON NON-NORMALIZED IMAGES AT DIFFERENT INTENSITIES BY USING CNN.

Intensity	45l	30l	15l	15r	30r	45r	Avg.
1	48.17	58.00	61.83	61.50	57.17	51.17	56.31
2	56.67	65.50	70.50	71.50	67.67	59.33	65.19
3	66.33	71.67	77.17	76.17	71.33	62.83	70.92
4	70.83	73.33	77.33	77.33	73.00	68.83	73.44
Avg.	60.50	67.13	71.71	71.62	67.29	60.54	**66.47**

database [27] in order to adapt it for facial expression recognition. We have further adapted it by fine-tuning it once more on the frontal training images from the BU3DFE database [18]. This procedure has been repeated for each fold and five different models have been obtained, *i.e. one for each fold*.

C. Results

The experimental results are given in Tables I-IV. Table I and Table II presents the results when using LBP+SVM based approach on the pose normalized and non-normalized face images, respectively. Similarly, Table III and Table IV convey results achieved with the CNN-based approach. In all the tables, the correct classification accuracies achieved at each intensity level are also given separately. It can be observed from Table I and Table II that pose normalization contributes significantly to the performance of the LBP+SVM based approach. The average accuracy has increased from 45.15% to 62.24%. This indicates that LBP features are very sensitive to pose variations and can benefit from pose normalization. As expected, as the pose angle increases, the performance drops in the cross-pose FER experiments. The highest improvement in the recognition accuracy is observed for 45 degree images, because it is more challenging to perform facial expression recognition when the pose angle difference between the training and testing images is high. However, increase in view angle also poses a challenge for the pose normalization method, since face frontalization quality deteriorates with increasing pose angle due to self occlusions. It can be seen that when the expression intensity increases the correct classification rates also increases. This is expected, since facial expressions are more apparent at higher intensities.

CNN-based facial expression recognition approach is found to be superior compared to the LBP+SVM based approach as shown in Tables III and IV. The average correct classification rate is increased from 45.15% to 66.47% when CNN-based approach is applied on the non-normalized images. It improves the average accuracy from 62.24% to 67.60% when it is performed on the pose normalized images. An interesting observation is that when CNN-based approach is used, the performance difference due to face frontalization is less. Compared to an absolute performance difference of around 17% in the case when LBP+SVM based approach is used, it is only 1.2% when CNN-based approach is applied. This outcome can be explained with the fact that, during the fine-tuning process of the VGG-Face model on the FER 2013 database, the CNN captures also the pose variations while learning to classify different expressions, since, as illustrated in Fig. 2, the FER 2013 database contains face images with varying pose angles. This indicates that CNN-based approach is robust to pose variations. However, pose normalization still contributes to the performance. A similar observation has been previously made in a CNN-based face recognition approach, where pose normalization has shown to contribute around 1-2% increase in face recognition accuracy [29]. Fig. 3 gives an overview of the results at different pose angles, comparing the average accuracies obtained both on the pose normalized face images and non-normalized 2D samples.

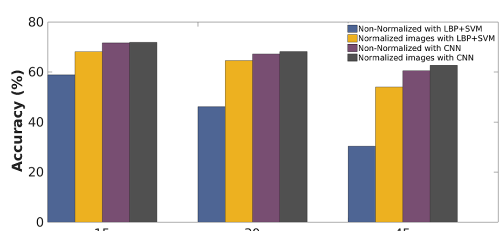

Fig. 3. Average recognition accuracies at different pose angles.

ISBN 978-90-827970-1-5 © EURASIP 2018

IV. CONCLUSIONS

In this paper, we have investigated the effect of pose normalization for cross-pose facial expression recognition. We have presented an expression preserving face frontalization method and employed two facial expression recognition approaches, one based on hand-crafted LBP features combined with SVM classifiers and the other based on a state-of-the-art deep CNN model, VGG-Face, which is fine-tuned on the FER 2013 database. We have conducted extensive experiments on the BU3DFE database using a cross-pose, subject-independent setup, in which frontal face images are used for training, whereas face images with varying pose angles, ranging from -45 to 45 degrees in 15 degrees steps, are used for testing.

In the experiments, CNN-based FER approach is found to be superior to the approach based on LBP+SVM combination. We have shown that face frontalization improves the performance significantly when LBP+SVM combination is used. However, the achieved improvement when CNN-based approach is employed is smaller. This shows that CNN-based approach can learn to handle pose variations up to some extent, since it is fine-tuned on a dataset that contains face images with varying pose angles. Another interesting observation is that the performance gain obtained in a CNN-based FER approach due to pose normalization matches the one that has been reported to be achieved by a CNN-based face recognition approach. This validates that, although CNN-based facial image analysis approaches are very powerful and can handle pose variations, when they are trained with samples from different view angles, one can still benefit from pose normalization to further improve their performance. We plan to continue to work on enhancing the quality of expression preserving face frontalization in order to achieve higher accuracies.

ACKNOWLEDGMENT

This work is partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 688900 (ADAS&ME project - http://www.adasandme.com)

REFERENCES

[1] H. Gao, H. K. Ekenel, and R. Stiefelhagen, "Pose normalization for local appearance-based face recognition," in *International Conference on Biometrics*. Springer, 2009, pp. 32–41.

[2] H. T. Ho and R. Chellappa, "Pose-invariant face recognition using markov random fields," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1573–1584, 2013.

[3] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Face frontalization for alignment and recognition," *arXiv preprint arXiv:1502.00852*, 2015.

[4] H. Gao, H. K. Ekenel, and R. Stiefelhagen, "Combining view-based pose normalization and feature transform for cross-pose face recognition," in *International Conference on Biometrics (ICB)*. IEEE, 2015, pp. 487–492.

[5] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust statistical face frontalization," in *ICCV*, 2015, pp. 3871–3879.

[6] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *CVPR*, 2015, pp. 4295–4304.

[7] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3d pose normalization," in *ICCV*. IEEE, 2011, pp. 937–944.

[8] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 980–993, 2015.

[9] C. Ding and D. Tao, "Pose-invariant face recognition with homography-based normalization," *Pattern Recognition*, vol. 66, pp. 144–152, 2017.

[10] W. Zheng and X. Zhou, "Cross-pose color facial expression recognition using transductive transfer linear discriminat analysis," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1935–1939.

[11] J. Hashemi, Q. Qiu, and G. Sapiro, "Cross-modality pose-invariant facial expression," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4007–4011.

[12] F. Guney, N. M. Arar, M. Fischer, and H. K. Ekenel, "Cross-pose facial expression recognition," in *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, pp. 1–6.

[13] O. Rudovic, I. Patras, and M. Pantic, "Facial expression invariant head pose normalization using gaussian process regression," in *CVPRW*. IEEE, 2010, pp. 28–33.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[18] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th international conference on*. IEEE, 2006, pp. 211–216.

[19] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3d Faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '99. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[20] D. Vlasic, M. Brand, H. Pfister, and J. Popovi, "Face transfer with multilinear models," in *ACM Transactions on Graphics (TOG)*, vol. 24. ACM, 2005, pp. 426–433.

[21] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: a 3d facial expression database for visual computing," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 3, pp. 413–425, 2014.

[22] V. Blanz, A. Mehl, T. Vetter, and H.-P. Seidel, "A statistical method for robust 3d surface reconstruction from sparse data," in *2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings*, Sep. 2004, pp. 293–300.

[23] N. Faggian, A. Paplinski, and J. Sherrah, "3d Morphable Model fitting from multiple views," in *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*, Sep. 2008, pp. 1–6.

[24] C. Cao, H. Wu, Y. Weng, T. Shao, and K. Zhou, "Real-time facial animation with image-based dynamic avatars," *ACM Transactions on Graphics*, vol. 35, pp. 1–12, Jul. 2016.

[25] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European Conference on Computer Vision*. Springer, 2004, pp. 469–481.

[26] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[27] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International Conference on Neural Information Processing*. Springer, 2013, pp. 117–124.

[28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[29] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *CVPR*, 2014, pp. 1701–1708.