A Novel Sample-Enhanced Dataset based on MFF for Large-Angle Face Recognition

He Wang[1], Yan Wang[1,*], Jie Liu[1], Guisheng Ying[1]
{wanghe11@hrbeu.edu.cn，wangyanj@hrbeu.edu.cn[*], liujie@hrbeu.edu.cn, yinguisheng@hrbeu.edu.cn }
[1] College of Computer Science and Technology, Harbin Engineering University, China

Abstract. Large-angle face recognition has always been a huge challenge due to the scarcity of large-angle dataset. In this paper, a novel sample-enhanced dataset is constructed, which is composed of various angle face picture samples from -90° to 90° relative to the front face. The constructed dataset is obtained by enhancing large-angle face samples of the CASIA-WebFace dataset. The large-angle face samples are generated from small-angle face samples of the CASIA-WebFace dataset, which is based on the multi-task feature framework (MFF). By employing these sample datasets, four trained FaceNets are achieved for face recognition. Finally, to test the effectiveness of the four face recognition networks for the large-angle face, 300 large-angle face pictures of different basketball players are selected as the samples of the experiment. The results demonstrate that the accuracy of large-angle face recognition has been greatly improved when utilizing the FaceNet that is trained by the novle enhanced dataset.

Keywords: Enhanced-dataset, large angle, MFF, face recognition.

1 Introduction

With the progress of technology, face recognition is greatly developed, which is applied everywhere in our daily life, such as the ticket checking system of railway station, the mobile phone face unlock system and so on. Most them are focused on small-angle face pictures. However, the angle of the face is various in different scenes. For the large-angle face pictures, the face recognition system may not work well. Yet, large-angle face recognition has always been a huge challenge due to the scarcity of large-angle dataset. In fact, large-angle face recognition plays a vital role in our daily life, such as identifying criminals in video surveillance. Therefore, constructing new large-angle face samples is particular important and essential, which will drive the progress of large-angle face.

With the development of neural networks, small-angle face recognition has been overcome. However, because it is difficult for neural networks to extract enough features for large-angle faces, high-angle face recognition has always been a huge challenge. At the same time, the published data sets such as CASIA-WebFace [1] and LFW [2] have rela-tively small angular deflection ranges, and there have been few reports of high-angle face recognition data sets in recent years, which has limited high-angle people Progress in face recognition. Although many people such as DR-GAN proposed by Liu et al. [3] and chu at al. [4] have proposed solutions for multi-pose face recognition, so far few people have tried solutions for large-angle face recognition from the perspective of datasets.

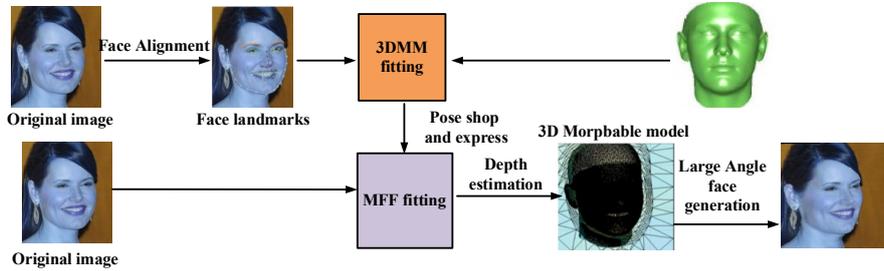In this paper, a 3D face reconstruction method based on the MFF [5] framework is used to

Fig. 1. Review large-angle face recognition sample generation method-3D face reconstruction based on MFF framework

generate a large-angle face sample set. It is the first time to use the MFF sample generation of 3D face reconstruction for the large-angle dataset application. MFF is a framework that can strongly constrain the pixel intensity and the 3DMM prior parameters to the BFM [6] face model. The overall of architecture of the paper is shown in figure 1. Firstly, we align the face samples to obtain Face landmarks, and then fit the Face landmarks to the 3DMM model to solve the shape, expression, pose, and lighting parameters required by the MFF algorithm. Finally, the face model reconstructed based on the MFF algorithm is subjected to depth estimation to complete three-dimensional reconstruction, and further generates face samples with larger angles. At last, based on the generated high-angle face sample set, the recognition of high-angle face images under the real environment of 15 basketball players is completed, which verifies the effectiveness of our sample generation method.

2 Method

In this section, the procedure of generating a multi-pose face dataset is introduced in detail, which includes face alignment, pose estimation, 3DMM coefficient solution and MFF frame generation of multi-pose faces.

2.1 Face alignment

Face alignment is one of essential and crucial work for face recognition. However, the womerk has become more difficult with the increasing of complexity of the background and the pose of face.

To imitate more realistic face recognition environment in our daily life, the CASIA-WebFace dataset [1] is employed in the experiment, which is obtained from a web crawler. The pictures from the CASIA-WebFace have more complicated lighting background to make it closer to the environment of face recognition in daily life. Based on the CASIA-WebFace dataset, the 3D face recognition method is utilized to generate a face recognition dataset with the performance of a large angle. Thus, the generated dataset can achieve the requirement of large angle and wild environment. For the face alignment of the pictures from the generated dataset that has more face poses and complicated environment, the face alignment network (FAN) method pro-posed by Adrian Bulat [7] is adopted, which is shown in figure 2. The adopted FAN method is constructed by stack-ing four Hour-Glass (HG) networks. The Hour-
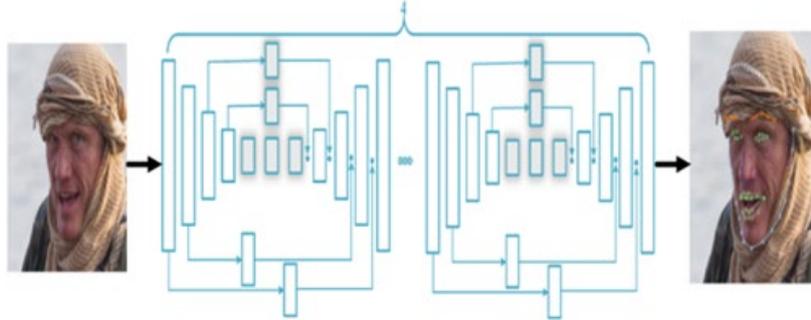
Fig. 2. The Face Alignment Network (FAN) constructed by stacking four HGs in which all bottleneck blocks. This network uses RGB images as input and face landmarks as output.

Fig. 3. Face alignment results of CASIA-WebFace using FAN method

Glass (HG) networks [8] are based on one of the state-of-the art architectures for human pose estimation. By doing the experiment .the final results are presented in figure 3. From figure 3, it can be concluded that FAN method is more robust for the complex background and lighting of the face in CASIA-WebFace and there are fewer failures.

2.2 Pose estimation

Pose estimation is an important part in the generation of large-angle face samples. The use of large-angle face sam-ples for projection to retain more identity information is one of the characteristics of 3D face reconstruction. We use 5 landmarks (nose tip, chin, left eye, left eye, right eye, (Left mouth corner, right mouth corner) as 2D face key points, 3D face key points, which are set based on the the work in [9]. The pose estimation of face images in CASIA-WebFace is completed. Due to the good effect of face alignment in the section 2.1, the accuracy of the pose estimation is improved. The results are shown in figure 4.

From figure 4, we performed pose estimation on the face samples in CASIA-WebFace. Where figure 4(a), (b), (c), and (d) are the cases of 45°, -45°, 56°, and -57° around the Y axis. In most cases, pose estimation can be performed accurately, but a lot of face occlusions will reduce the accuracy of the pose estimation method.

Fig. 4. Face pose estimation effect map.

2.3 3D face model fitting

We use the 3DMM fitting method to obtain the 3DMM prior parameters that are needed for the MFF algorithm. The 3DMM model is a universal 3D face model [10]. It can approach any real face shape under the constraints of shape and texture conditions. We refer to the work of Chu et al. [4] to solve 3DMM parameters.

$$S_{i,e} = \left(X_1, Y_1, Z_1, X_2 \cdots Y_{n_{vert}}, Z_{n_{vert}}\right) \tag{1}$$

Each 3D face shape (with n_{vert}) is represented by a $3n_{vert} \times 1$ dimensional vector, which stands for (0-6) seven expressions, $i \in (0, n_{id} - 1)$ for the index value of identity.

$$S = \bar{S} + A_{id}\alpha_{id} + A_{exp}\alpha_{exp} \tag{2}$$

Where S is the 3D face, $\bar{S}$ is the average shape. We use BFM, which is the most widely used 3D spatial face model. A_{id} is an identity morphable model whose principal axis of variations are in the matrix. α_{id} is the coefficient. A_{exp} is the principle axes trained on the offset between expression scans and neutral scans. α_{exp} is the weight.
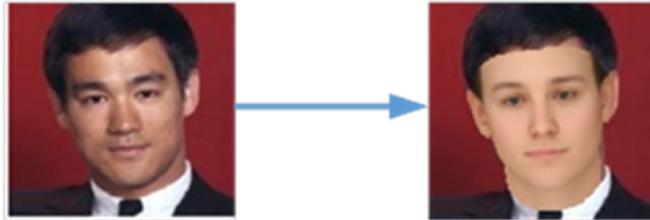

Fig. 5. BFM model and 3DMM parameter fitting.

Finally, formula (3) was used for shape constraint, requiring the key points of the target 3D model to coincide with the key points of the 2D image after being projected.

$$arg_{f,R,t_{3d},\alpha_{id}\alpha_{exp}} = min\|S_{2dt} - S_{2d}\| \tag{3}$$

Where f is the scale factor, and R is the rotation matrix. Finally, we follow the work of liu at al [9]. The PCA dimension reduction method is used to solve the 3DMM prior parameters required by the MFF framework. The effect of BFM model and parameter fitting is shown in figure 5..

2.4 MFF framework and generate large angle face

MFF is a multi-feature task framework that can reconstruct the 3D face with identity information under the constraints of 3DMM prior parameters and texture conditions, which is presented in figure 6.
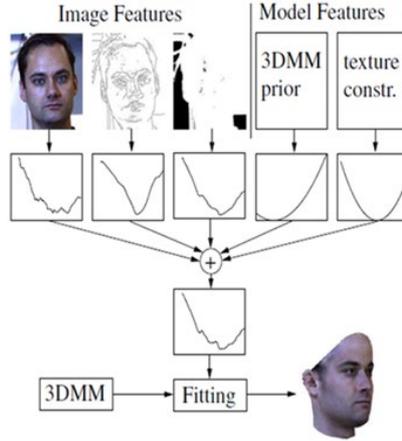

Fig. 6. Schematic diagram of multi-tasking framework MFF (Pixel inten-sity feature, Edge feature, Specular highlight feature, Gaussian Prior feature, Texture constraint feature)

The superiority of the MFF framework is reflected in the fact that he constrains 5 features (Pixel intensity feature, Edge feature, Specular highlight feature, Gaussian Prior feature, Texture constraint feature) at the same time, and there are few cases of failure. The purpose of the fitting algorithm using multiple features is to maximize the posterior probability of the model parameters, not only as done in SNO [11], but also to consider the different characteristics of the model parameters. Its fusion algorithm is shown in formula four:

$$\min_{\theta} \tau^c C^c + \tau^e C^e + \tau^s C^s + \tau^p C^p + \tau^t C^t \tag{4}$$

In this function, the C^c is the pixel intensity of the picture, C^e is the edge feature, C^s is the specular highlight feature, C^p is the 3DMM prior parameters, finally, C^t is the texture constraints. τ^s is the feature weighting factors. Among them, Pixel intensity feature, Edge feature, Specular highlight feature, three features are obtained from the input RGB image, and

Gaussian Prior feature is obtained from the 3DMM prior parameter of 2.2. It is minimized using a Levenberg-Marquardt optimization algorithm [12].

Fig. 7. Large-angle face generation results (a) 30° (b) 60° (c) 75° (d) 90°

After using MFF multitasking to form a strong constraint on facial identity information, we follow the method of Liu et al. To mark some anchor points outside the surface area and estimate their depth. After the depth information is obtained, a large-angle face image can be generated. As shown in Figure 7. Among them, (a) is the input picture, and the yaw angle is estimated to be 30 degrees after attitude estimation. (b), (c), and (d) are the results after the large angle is generated, and the identity information is better retained.

3 Experiments

In this section, the experiment is implemented to verify the effectiveness of the enhanced large angle face datasets, which is divided into two parts. In part 1, the generation of the large angle face dataset is introduced in detail. The results of the test of the generated large angle face dataset are analyzed and discussed in part 2, which is a good validation for the generated large angle face dataset.

3.1 Large-angle face dataset generation

3.1.1 Datasets

To generate and test the large angle face dataset, another two datasets, as the fundamental samples, are employed in the experiment. The first one is the CASIA-WebFace dataset, which is composed of 10575 subjects and 494414 images. The pictures from the CASIA-WebFace dataset are produced from the Internet. Thus, the sample pictures of the CASIA-WebFace dataset have more complicated background, which can be used to be closer to the environment in our daily life.

The second dataset we used is composed of several real pictures, which is to verify the effectiveness of the method generating the sample set. In this dataset, face pictures of 15 basketball players are extracted. These face pictures have various angles, most of which have yaw angle greater than 60°around the y-axis. The backgrounds of these face pictures are from different basketball game, which can keep these face pictures in a wild state. The sample dataset is constructed by 300 face pictures in total and 20 face samples are extracted from every basketball player in average. Each image of these face pictures is cropped to 160*160 px. The constructed dataset is named as validation set and some of the face picture samples are presented in figure 8.

Fig. 8. Validation data set, the first row is LeBron James, the second row is Kevin Durant.

3.1.2 Sample set generation

To generate high angle face samples, 100 subjects are extracted from the CASIA-WebFace dataset, which is called CASIA-WebFace (100). The face alignment of these selected 100 subjects is finished by using the method of FAN. Then the selected dataset is divided into two intervals (-90°-0° and 0°-90°) by employing the method of pose estimation using five key points mentioned in the section 2.2. To more accurately estimate the rotation angle of faces pictures, the divided two big intervals are further partitioned into smaller intervals by the step of 15°. Finally 12 sample intervals are obtained, which is shown in table 1. From table 1, it can be concluded that most of the samples in the CASIA-WebFace (100) are focused on the interval from -60°-60° and a few samples are distributed outside the interval from -60°-60°. Therefore, some additional work is conducted to enhance the samples especially in the interval from -90°-60° and 60°-90°.

Fig. 9. Some samples in Sample-Data.

The detail procedure of sample enhancement has been mentioned in the section 2.3 and 2.4. Firstly, 3DMM Fitting method is adopted to solve the prior parameters of the former CASIA-WebFace (100). These prior parameters are required by the MFF framework. Based on the MFF 3D face reconstruction, the lower angle face pictures in CASIA-WebFace (100) can be well projected to higher angle face pictures. The corresponding intervals are detail depicted as flowing:

- Projecting samples in (30°, 45°) (-45°, -30°) in CASIA-WebFace(100) to (46°, 60°) (61°, 75°) (76°, 90°) (-60°, -46°) (-75°, -61°) (-90°, -76°) within six intervals.

- Projecting images in (46°, 60°) and (-60°, -46°) sections of CASIA-WebFace (100) to (61°, 75°), (76°, 90°) (-75°, -61°), (-90°, -76°) within 4 intervals.

Finally, the projected samples are added into the original CASIA-WebFace (100) and a new sample dataset can be obtained. The newly generated sample (it is called as Samples-Data) have many face sample no matter any intervals from -90° to 90°. The detail comparison of both samples between the original CASIA-WebFace (100) and Samples Data is presented in table 1.

Table 1. The comparison of the CASIA-WebFace (100) dataset and Samples-Data

Angle	CASIA-WebFace(100)	Samples-Data
[-90°~-75°)	0	818
[-75°~-60°)	16	834
[-60°~-45°)	229	818
[-45°~-30°)	589	589
[-30°~-15°)	806	806
[-15°~-0°)	1089	1089
[0°~-15°)	996	996
[15°~30°)	865	865
[30°~45°)	608	608
[45°~60°)	225	833
[60°~75°)	12	845
[75°~90°)	1	834

By the comparison in the table 1, it can be obviously deduced that the number of the samples from the CASIA-WebFace (100) dataset has greatly increased in the intervals from -90°--45° and 45°-90°. Especially in the higher intervals from -90°--60° and 60°-90°, the growth is the most pronounced and the number of the samples have increased at last 818, which has greatly expanded the CASIA-WebFace (100) dataset in the aspect of large angle face dataset. Also, the large angle face recognition is driven by the generated dataset Samples-Data.

We generated a large-angle face recognition sample set of 12 angle intervals, as shown in Figure 8. The first line is -90°-0°, and the second line is 0°-90°.

3.2 Validation of large-angle face generation method

Most of the face recognition methods are relied on the deep learning. Transfer learning [13] is a hot research topic in the field of deep learning, which can increase the efficiency of training. The expanded dataset Samples-Data is developed by the CASIA-WebFace (100), which mainly focuses on the large angles of the CASIA-WebFace (100). The FaceNet pre-trained models [14][15] for CASIA-WebFace transfer learning training, and only trained the last softmax layer to extract high-level features.

The Samples-Data sample set is divided into 3 intervals, which are $0°<=|Y|<=60°$, $0°<=|Y|<=75°$, $0°<=|Y|<=90°$ (Y is around the Y-axis Rotational yaw angle) respectively. The FaceNet is used to train the sample sets in three intervals and CASIA-WebFace (100) to investigate the influence of different face angles on the extracted features of the neural network, which can be used to further verify the effectiveness of the sample enhancement method. FaceNet is trained using four types of sample sets and the four neural networks after fine-tuning training are named A, B, C, and D, respectively. The validation set consisting of 300 basketball

player face pictures are utilized to verify the effect of large angle face recognition. The results are shown in table 2.

Table 2. FaceNet recognition accuracy of the validation set under four dif-ferent training sets.

Training	A	B	C	D
Accuracy	0.8026	0.8438	0.8672	0.8012

From table 2, the minimum recognition rate of the validation set using A network is 0.8026 and D network is 0.8012. Because the training sets of A and D lack large angle face picture samples. The recognition rate of B network is 0.8438, which is greatly improved compared to A and D. Because the training set of $60°<= |Y|<=70°$ is added to make the neural network learn more deep features. The recognition rate of the verification set under the C network reaches 0.8672, which is the highest accuracy of the four networks. The main reason is that the training set of the D network includes faces in extreme states (the face has a yaw angle of 90 around the Y axis). The samples allow the neural network to learn enough features.

4   Conclusion

In this paper, a novel sample-enhanced dataset is constructed, which is composed of various angle face picture samples from -90° to 90° relative to the front face. To generate the constructed dataset, multi-task feature (MFF) framework is employed in this paper. The generated dataset is developed by the CASIA-WebFace dataset, which greatly expands the samples of the CASIA-WebFace dataset especially for the large angle samples. The FaceNet is employed and trained by using the generated dataset. To test the effectiveness of large-angle face recognition, 300 large-angle face pictures of different basketball players are selected as another sample by using four different trained networks. The results demonstrate that the accuracy of face recognition has been greatly improved when the larger angle face picture samples are used to the network.

References

[1]   Yi D, Lei Z, Liao S, et al. Learning face representation from scratch, *Computer Science*, arXiv:1411.7923v1 [cs.CV], 28 Nov., 2014.
[2]   http://vis-www.cs.umass.edu/lfw/
[3]   Tran L, Yin X, Liu X. Disentangled representation learning GAN for pose-invariant face recognition, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 21-26 July, Honolulu, HI, USA, 2017.
[4]   Chu B, Romdhani S, Chen L. 3D-aided face recognition robust to expression and pose variations. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1907-1914, 23-28 June, Columbus, OH, USA, 2014.
[5]   Romdhani S, Vetter T. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 20-25 June, San Diego, CA, USA, USA, 2005.
[6]   Paysanp, P., Knothe R., Ambegrg B., et al. A 3D face model for pose and illumination invariant face recognition. *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2-4 Sept, Genova, Italy, 2009.

[7] Bulat A., Tzimiropoulos G., How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). *2017 IEEE International Conference on Computer Vision (ICCV)*. 22-29 Oct., Venice, Italy, 2017.

[8] Newell A., Yang K., Deng J.. Stacked hourglass networks for human pose estimation. arXiv:1603.06937v2 [cs.CV], 26 Jul, 2016.

[9] Zhu X., Lei Z., Yan J., et al. High-fidelity pose and expression normalization for face recognition in the wild. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 7-12, June, Boston, MA, USA, 2015.

[10] Blanz V., Vetter T., Rockwood A.. A morphable model for the synthesis of 3D face. *Acm Siggraph*, pp. 187-194, 2002.

[11] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 1063 – 1074, Sept., 2003.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. Numerical recipes in C : the art of scientific computing. Cambridge University Press, 1992.

[13] Pan, Sinno Jialin, Yang, Qiang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge & Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

[14] Schroff F., Kalenichenko D., Philbin J., FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 7-12, June, Boston, MA, USA, 2015.

[15] https://github.com/davidsandberg/facenet