PAPER • OPEN ACCESS

Normalization of Facial Pose and Expression to Increase the Accuracy of Face Recognition System

To cite this article: Muhammad Djamaluddin *et al* 2020 *J. Phys.: Conf. Ser.* **1539** 012035

View the article online for updates and enhancements.

Normalization of Facial Pose and Expression to Increase the Accuracy of Face Recognition System

Muhammad Djamaluddin[1], Martua Hamonangan N[1], Aswian Editri S[1]

[1]Engineering Faculty, Hamzanwadi University, West Nusa Tenggara, Indonesia

Abstract. Normalization of face poses and expressions is one of the critical steps to restore the raw appearance of faces in uncontrolled conditions (in the wild) and help improving the performance of face recognition systems. Researchers in the field of computer vision has for a long time trying to find methods to normalize of poses and expressions to gain accuracy improvement for the face recognition process. The process of normalization is related to the face alignment stage, which aims to find face landmarks (point markers of facial features). From this face landmarks, the poses and facial expressions that are different from canonical face obtained for later transformation. This paper discusses the method of normalizing poses and facial expressions that have been proposed by researchers from geometrical-based transformation to the current state of the art, which is a deep learning-based system.

1. Introduction

Face Recognition (FR) is one of bio-metric methods in addition to fingerprints and eye corneas which since the early 1970s has become a hot research topic that has evolved to the present. The topic is fascinating to study because of its non-obtrusive nature. People do not need to be in close distance to the camera sensor for the system to perform recognition as in case of fingerprints or corneas. The potential application of FR research is quite a lot, including crime countermeasures, securing critical facilities, securing information systems and multimedia systems. Face recognition in the wild is still a challenge because so many environmental factors affect the quality of facial images that would be detected during camera acquisition such as differences in face poses, lighting, camera resolution or occlusion of the face.

There are 2 typical applications of face recognition, Identification and Verification. Identification means given a picture containing a face; the system would find the identity of the person who has that face. The purpose of verification is to verify whether a face image matches with the person who wants to be verified. These types of application if applied in the wild or an uncontrolled environment, often face obstacles in terms of robustness and performance accuracy. Moreover, since the end-to-end FR system consists of several stages, improvements in each stage can contribute to improving the accuracy and robustness of the system in general. Face Recognition (FR) is one of the bio-metric methods in addition to fingerprints and corneas, which since the early 1970s has become a research topic that has evolved to the present. The topic is fascinating to study because of its non-obtrusive nature. People do not need to be in close distance to the camera sensor for the system to perform recognition as in case of fingerprints or corneas. While it still raises some ethical questions, the potential application of FR research is enormous, including crime countermeasures, securing critical facilities, securing information systems, and multimedia systems. Face recognition in the wild is still a challenge because so many environmental factors affect the quality of facial images that would be detected, such as differences in face poses, lighting, camera resolution or face occlusion.

There are 4 steps in the face recognition algorithm process, as shown in Figure 1:

1. **Face Detection**: Detects and finds faces in an image and marks their location.
2. **Face Alignment:** For each face image found, find facial landmarks in the face that are useful for normalizing non-standard face images in the face recognition process.

3. **Feature Extraction/Representation:** Find unique features of aligned-faces that distinguishable with others and represent them in a vector quantity.
4. **Face Matching/Recognition:** Compares the vector quantity of features obtained with a database containing a collection of feature vectors to perform the recognition process.

Figure 1. Four Stages of Face Recognition System

At present, the performance accuracy of the FR system for specific standard datasets for face in the wild such as LFW datasets is very high and resembles human capability. However, when implemented on other more massive datasets with a wide variety of poses, accuracy decreases dramatically [1]. Especially if the system is implemented directly in real-time, the accuracy generated is still far from human ability. One reason is the enormous variation in environment, media acquisition limitation, variation in poses, facial expressions, and occlusion.

One technique to increase the accuracy of Face Recognition is to normalize poses and facial expressions after the face patch detected. This normalization process is performed following Face Detection and Face Alignment stages, where Face Detection finds the location of faces in the picture/frame, and Face Alignment looks for facial landmarks in the face. Just before facial features extracted following the discovering facial landmarks, the normalization techniques performed. This paper address some of the pose and expression normalization techniques that have been developed today with the ultimate goal of helping to improve the accuracy of face recognition systems.

2. Face Alignment

(a) Pose   (b) Occlusion   (c) Expression   (d) Illumination

Figure 2. Problem in non-ideal face image [2]

Face Alignment tries to find facial landmark or fiducial facial points in a face. In a limited environment it is considered by many a solved problem, meaning that its accuracy is very close to ground-truth, but in an unlimited environment it is still an open and difficult problem. The reason is the high degree of variation of face appearances in this environment caused by dynamic traits that exist in the face or changes in the environment. There are 4 main factors that affect face alignment in the wild as shown in Figure 2:

1. *Pose,* the appearance of local facial features is very diverse because the face pose variation to camera (frontal, upside down etc.).
2. **Occlusion**, face partially covered for example by hair, glasses
3. *Expression*, some parts of the face such as mouth and eyes are very sensitive to differences in expression
4. *Illumination*, lighting is very influential on the overall appearance of the face

The main aspect of face alignment problem is to look for a set of facial points / face landmarks in the face image. The number of facial points varied, from minimum 5 points to 88 points. Usually this process starts from a simple form of facial point model and then processed iteratively for improving the estimated model shape until it reaches to an optimal form.

Facial Alignment methods can be divided into 2 categories, namely:

1 *Generative Methods*: Construct generative models for face shape and appearance. Face alignment problem is considered as an optimization problem to find the optimal form display parameters in producing a model that matches the test data. The best-known generative face alignment method is the Active Appearance Model (AAM) and its derivatives. AAM consists of 3 components: Shape Model, Appearance Model, and Motion Model. The Shape Model represents face shape as a set of k number of points where k can be as small as 5 or standard 68. To learn the shape model, the face shape is trained and normalized using global similarity transform (such as Procrustes analysis). The *Principle Component Analysis* (PCA) method is then applied both to the parameters of the Shape and the Appearance Model.

2 *Discriminative Methods*: These methods directly resolve the target face landmark location from facial appearance through methods such as learning local detector on each face landmark or learning regression function to look for general facial shapes. The approach in the discriminative method is to examine discriminatory functions that perform direct mapping between face images and facial landmark models. Examples are Cascaded Linear Regression, Local Constraint Model, and others. Before Deep learning era, the cascade regression method was state of the art for face alignment.

Some recent deep learning studies for face alignment combine Convolutional Neural Network (CNN) with methods such as cascaded regression [4], heatmap [5] [8], recurrent regression [6], the combination of heatmap and cascaded regression [9]. Recurrent Neural Network (RNN), especially one of its variants, namely Long Short-Term Memory (LSTM) is also used to perform face alignment by combining it with CNN for feature extraction [7]. The combination of RCNN and CNN referred to as Convolutional Recurrent Neural Network (CRNN).

Adrian et al. [8] proposed a Deep Neural Network called the Face Alignment Network (FAN) based on one of the states of the art architectures for estimating human poses, namely the *HourGlass* (HG) network proposed by Newel et al. [10]. FAN was built by stacking four *HourGlass* Networks in which all bottleneck blocks (described as rectangles) were replaced with hierarchical, parallel, and multi-scale blocks from Binarized Convolutional Landmark Localizers used by Adrian et al. [11]. This block outperforms the Residual Block on the Residual Neural Network for image classification. Area Under Curve (AUC) is used to calculate the performance of FAN architecture, with a 7% threshold on several primary 2D face alignment datasets such as MDM, CFSS, and TCDCN using the ground truth boundary box. Except for Category C 300-VW, it is evident that 2D-FAN achieves the same performance on all datasets, outperforming MDM and ICCR, and similar to MDM-on-LFPW performance level. Since 2D-FAN resembles MDM-on-LFPW performance, it signifies that 2D-FAN achieves almost saturated performance in the 2D dataset.

Marek et al. [12] introduced the Deep Alignment Network (DAN) which was inspired by the Cascade Shape Regression (CSR) framework, and like CSR this method begins with the initial estimation of facial shape (S_0) which is refined through several iterations. However, in DAN, each CSR iteration is replaced by a single stage of Deep Neural Network that performs feature extraction and regression. The main difference between DAN and the CSR approach is that DAN extracts features from all face images rather than just patches/faces around a landmark location. Each stage of the Neural Network in DAN fixes the estimated location of landmarks produced by the previous stage, starting with the initial estimation So. The inter-layer connections build a link between successive stages of the network by producing heatmap landmarks H_t, Image Features F_t, and transformations T_t to adjust and correct input images to standard poses. By introducing the landmark heatmap and drawing features, important information including the approximate location of the landmark can be transferred between

stages. Extensive evaluation of two challenging and publicly available datasets shows that DAN decreases the state-of-the-art failure rate with a significant margin of more than 70% so that the mean error alignment is drastically reduced.

From the two examples of Deep Neural Network configuration above, it can be concluded that the performance of the face alignment by calculating the reduction of mean error, has increased quite dramatically compared to traditional methods, both generative and discriminative methods so that the DNN-based method becomes state of the art for face alignment.

3. Face Pose and Expression Normalization

Several studies show the importance of the normalization poses, expressions, and lighting process when performing face recognition because unideal factors cause intra-personal variations higher than inter-personal variations. People with different poses and expressions will be more difficult to recognize than recognizing some individuals with a standard pose.

The problem of poses normalization has widely been studied. For example, Chai et al. [14] in 2003 already proposed the use of *Affine Transformation* to normalize a nearly frontal face pose where the angle of distortion is between 40 degrees to the left and right of the standard pose. Broadly speaking the method is as follows: divide the face area into three rectangles separated by positions of iris. By studying the relationship between a particular pose and a frontal pose, the affine transformation's parameters can be calculated to normalize face geometrically into a frontal pose. Chai et al. prove that with this form of normalization, face recognition's accuracy increases by a maximum of 24% for certain pose angles. However, the method has many limitations, especially the allowed pose is limited to only a few types of angles from the "yaw" coordinates, namely 15, 25 and 40 degrees to the left and right. Besides, pose determination was based on the iris location, which implies the face position must be relatively closed to the camera sensor.

In addition to 2D image representations, 3D models are also used to overcome the problem unideal poses and expressions since poses variation captured in the image is actually due to the rigid transformation of the 3D face, so that the 3D model naturally more accurate and intuitive. For the 3D model, the 3D Morphable Model (3DMM), a 3D face model developed by Blanz et al. [15] commonly used for face modeling.

The 3DMM model is conventionally constructed using a statistical approach to a large 3D face scan dataset (geometric and texture data) and exerts domain knowledge about face variations using the pattern classification method. The Morphable Face Model is a multidimensional 3D morphing function based on a linear combination of a large number of 3D face scans. By calculating the mean face and the main variation modes in the dataset, the probability distribution is imposed on the morphing function to avoid faces that might not exist in the real world. It is a similar concept to the eigenface but in 3D face model dimensions [15]. From constructing a parametric face model that capable of producing almost all faces, the problem is modifiable to be an optimization problem of mathematical functions. New face images or 3D face scans, can be registered by minimizing the difference between a new face and the results of its reconstruction by the face model function. Recently Luan et al. [16] proposed a 3DMM reconstructing algorithm using 2D images instead of 2D images and 3D scans such as the original method proposed by Blanz. Another benefit of this method while the Blanz method utilizes PCA to decrease the vector dimensions of face shape and textures that can cause loss of non-linear features in the process. Luan et al. applied Deep Learning to reconstruct the 3DMM Model from a set of 2D face images.

Zhu et al. [18] used 3DMM to normalize poses and expressions by combining them with facial landmarks of a face image. This method is called High-Fidelity Pose and Expression Normalization (HPEN). In this method, face landmarks were matched and combined with a standard 3DMM model in

term of pose, shape, and expression. After the 3D model with poses and expressions resembling the input image was formed, then poses and expressions normalization is carried out by rendering the transformed 3D model with matching texture and shape. When transforming 3D models into standard 2D face shapes, there would be a missing part of the face image. HPEN would fill the missing pixel part with Poisson editing algorithm. The final result is a frontal face image with a standard expression, which ideal for the next stage of Face Recognition. Experimental results showed that HPEN improves face recognition accuracy performance by 1.4% and 2.03% using HD-LBP and HD-Gabor detection, where the combination of HPEN, HD-Gabor, and Joint Bayesian achieves state of the art results.

Attempts to synthesize normal face from a face image that invariant in poses, expressions, and occlusions are proposed by Forrester et al. [18] in CVPR 2017. In this method, the image features as an output from the face recognition process (in the paper using either FaceNet or VGGNet) would be decoded back into a normal face. FaceNet is a face recognition system based on Deep CNN whose fully connected layer produces a vector of 128 dimensions. For this face normalization, the feature vector produced by FaceNet was expanded to 1024. Hence, the Forrester et al. method re-mapped facial identity features to the shape of the face image. The main idea is to exploit invariant facial identity features for poses, lighting, and expressions by transforming the problem into a problem of mapping feature vector into face shapes with bright lighting, neutral and forward-facing expressions, commonly called normal faces. The facial identity feature is very reliable as discriminatory, so the network decoder is very robust against various interfering factors such as occlusion, lighting, and pose variations.

4. Research Opportunity

For the topic of pose and expression normalization, several states of the art techniques such as HPEN [18], Forester et al. [19] show robust normalization results on variations in expression and pose, but the quality of the output is still not optimal. HPEN [18] produces a 'damaged' and unnatural face image output, [19] produces a natural but blurry face image. There are also new techniques to normalize pose and expression in deep feature space level, instead of in image level as in HPEN, Affine Transformation etc. So, there are still rooms for further research in this face normalization topics.

In addition to variations in poses and expressions, occlusion problems that cause some faces covered with other objects such as hair, glasses, veil are also problems with a lot of research interest. Most papers examine the problem of normalizing variations in poses and expressions, but normalization of facial occlusions just started to gain interest. Research [19] that uses Generative Adversarial Networks (GAN) to produce facial images, for example, can be a baseline for further research

5. Conclusion

Normalizing facial pose and expression is one of the essential steps for increasing the accuracy of face recognition. Experiments conducted by researchers on public datasets showed that face recognition performance improved after normalization of unideal face images compared to not being normalized. Research on the normalization of poses and expressions is quite active topic where there are rooms for improvement.

6. References

[1]   Adrienne Lafrance 2016 The Ultimate Facial-Recognition Algorithm, *The Atlantic*, 28 Juni 2016, https://www.theatlantic.com/technology/archive/2016/06/machine-face/488969/

[2]   Yue Wu, Qiang Ji 2019 Facial Landmark Detection: A Literature Survey *International Journal of Computer Vision* vol **127** pp 115–142

[3]   V Kazemi, J Sullivan 2014 One Millisecond Face Alignment with an Ensemble of Regression Trees *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2014)

[4]   Hanjian Lai, Shengtao Xiao, Zhen Cui, Yan Pan, Chunyan Xu, Shuicheng Yan, Deep Cascaded Regression for Face Alignment, https://128.84.21.199/abs/1510.09083v1

[5]   Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski 2017 Deep Alignment Network: A convolutional neural network for robust face *Proceeding IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW)

[6]   Byung-Hwa Park, Se-Young Oh, Ig-Jae Kim 2017 Face alignment using a deep neural network with local feature learning and recurrent regression *Expert Systems with Applications: An International Journal* Volume **89** pp 66-80

[7]   Qiqi Hou, Jinjun Wang∗, Ruibin Bai, Sanping Zhou, Yihong Gong, 2018 Face alignment recurrent network *Journal Pattern Recognition* Volume **74** Issue C pp 448-458

[8]   Adrian Bulat and Georgios Tzimiropoulos 2017 How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks), *Proceeding ICCV2017*

[9]   Shahar Mahpod, Rig Das, Emanuele Maiorana, Yosi Keller, and Patrizio Campisi 2019 Facial Landmark Point Localization using coarse-to-Fine Deep Recurrent Neural Network https://arxiv.org/abs/1805.01760

[10]   A. Newell, K. Yang, and J. Deng 2016 Stacked hourglass networks for human pose estimation *Proceeding ECCV 2016*

[11]   A. Bulat and G. Tzimiropoulos 2017 Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources *Proceeding ICCV 2017*

[12]   Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski 2017 Deep Alignment Network: A convolutional neural network for robust face Alignment *Proceeding 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE*

[13]   D. Chen, X. Cao, F. Wen, and J. Sun 2013 Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification *Proceeding Computer Vision and Pattern Recognition (CVPR), 2013* pp 3025–3032

[14]   Xiujuan Chai, Shiguang Shan, Wen Gaol 2003 Pose Normalization for Robust Face Recognition Based on Statistical Affine Transformation *Proceedings of the 2003 Joint Fourth International Conference on Information, Communications and Signal Processing,* 2003 *and the Fourth Pacific Rim Conference on Multimedia*.

[15]   Volker Blanz, Thomas Vetter 1999 A morphable model for the synthesis of 3D faces *Proceeding SIGGRAPH '99 Proceedings of the 26th annual conference on Computer graphics and interactive techniques* pp 187-194

[16]   Luan Tran, Xiaoming Liu 2018 Nonlinear 3D Face Morphable Model *Proceeding of IEEE Computer Vision and Pattern Recognition (CVPR 2018) IEEE*

[17]   Liu Ding, Xiaoqing Ding, Chi Fang 2012 Continuous Pose Normalization for Pose-Robust Face Recognition *IEEE Signal Processing Letter, Vol.* **19**, No. 11

[18]   Zhu et.al 2015 High-Fidelity Pose and Expression Normalization for Face Recognition in the wild, Proceeding of *IEEE Computer Vision and Pattern Recognition* (CVPR) 2015

[19]   Forrester Cole, David Belanger, Dilip Krishnan, Aaron Sarna, Inbar Mosseri, William T. Freeman, 2017 Synthesizing Normalized Faces from Facial Identity Features *Proceeding of IEEE Computer Vision and Pattern Recognition (CVPR 2017)*