Pose and Expression Normalization for Face Recognition Using Delaunay Triangulation

Muhammad Yusuf Fadhlan[1,*] Dianthika Puteri Andini[1]

[1]*Department of Electrical Engineering, Bandung State Polytechnic, Bandung, Indonesia*
[*]*Corresponding author. Email: muhammad.yusuf.ttel410@polban.ac.id*

ABSTRACT

Detecting face image with various poses and expressions is a challenging problem especially if the face database consists of only neutral faces. To solve that problem, the pose and expression of face image should be neutralized first. This study proposes a method to recognize the face image by making the input image to be neutral in pose and expression to make it similar to the face images in the virtual database. The frontal face image needs to be neutralized for the expression using Delaunay Triangulation face model. For correcting pose view, a mirroring method is included in the transformation process. The final image that has obtained is evaluated by comparing it to the virtual face database. The face database; namely; Tarrlab face database is used for the experiment. Local Binary Pattern is chosen for the face recognition method. On average, this proposed study has been achieved improvement in recognition accuracy by 47.34%. A significant improvement can be shown at input images that have the extreme angle pose like 30 and 45 degrees.

Keywords: *Face Recognition, Expression Normalization, Pose Normalization, Delaunay Triangulation, Neutral Face Reference.*

1. INTRODUCTION

In a security system, face recognition is one of the fundamental functions. Face recognition also plays a crucial role in many implementations such as person identification systems, surveillance, human-computer interactions, etc. Face detection of an image is the important step in face recognition. In consequence, the efficiency and accuracy of face detection affects the performance of face recognition. Face recognition system is usually applied to identify a person from the face image with a neutral face database. For example, passport, ID-card or MUG shot for police records. However, there are many challenges occur in recognizing faces. The variations of input image due to pose, illumination, occlusion, and expression become one of the causes of face recognition failure. Among them, pose and expression have always been challenges. To solve the problem of pose and expression, many studies have been developed to normalize the expression and get the frontal face image. Ramachandran et al. [1] studied an improvement of the accuracy for face recognizing with the neutral database by using Candide face model and piece wise affine warping. Yet, this method takes expensive computational time though it still can improve the accuracy of recognition with principal component analysis (PCA) and linear discriminant analysis (LDA) method. Then, Hsieh et al. [2] used optical flow to warp the neutral face by synthesizing lip from the input images. The accuracy of face recognition can be improved by using that method but the optical flow method still has problems with the computational time. Later, Chayanut et al. [3] have applied a method to solve the computational time problem. They simplified the method by using a single face reference method. The single face reference is used as a template for transforming various (expression and pose) input image into a neutral face image. Yu-Jin Hong et al. [4] proposed a 3D based facial expression synthesis and pose frontalization. Three-dimensional facial reconstruction technologies are used in this research. B. Chu et. Al [5] also using 3D model for neutralize expression and pose variations. 3D Morphable Model methods is applied to represent the space of the human face. However, expression robust face recognition has been stealing attention many researchers [11] [12] [13] [14]. This requires the availability of face database with systematic expression variations data set such as the AR dataset [14], PIE [19] and Multi-PIE [10].

Yet, the pose of the input image still become a problem for some method that is not using 3D model. Pose correction for frontal image still looks unnatural especially for extreme degree between 30 degrees and

45 degrees. The 3D model can successfully solve this problem but computational time is still become a problem.

1.1. Objective

In this research, a better method for reaching a better accuracy of face recognition is proposed. Not only the better accuracy but also the better computational time is obtained. This research avoids using 3D model to realize the system to achieve high computational time. This proposed method is divided into two steps. First, the transformation method is selected to solve the expression problem. Secondly, to get frontal face image, the mirroring method is applied. Lastly, the final output image that has obtained is evaluated by comparing it to the virtual face database. Local binary pattern (LBP) face recognition method is chosen for the method. This proposed study improves the face recognition accuracy to recognize the image from virtual natural database especially for recognizing the input face image that has the extreme angle pose.

1.2. Paper Structure

This paper is divided into four main chapters. Chapter 1 provides the introduction of the research and the objective. Chapter 2 gives some background information about face recognition and expression normalization. Chapter 3 explains expression neutralization method using transformation method based on single reference model and pose normalization using mirroring technique. In chapter 4, the experimental results are presented. Also, conclusion and recommendation conclude the paper and presents direction for future research.

2. BACKGROUND

2.1. Face Detection

Face detection went mainstream in the early 2000's. A way to detect faces is invented by Paula Viola and Michael Jones [6]. However, many better solutions exist now. In 2005, a method called Histogram of Oriented Gradients is developed [7]. The concept of this face detection, the method will compute every single pixel in the image at a time. In every single picture, it will look at the pixels that directly surrounding it. This method compares how dark the current pixel compared to the surrounding pixel. Then an arrow is drawn that shows in which direction the image is getting darker.

2.2. Facial Landmark Estimation

Humans can easily recognize the similarity between two images but the computer would sometimes see the same pictures (same person) as two completely different people, especially when the pose and expression are totally different. To solve this problem, the computer needs to warp the picture so that the eye and lips of the images are always in the sample place of the image. By doing this process, comparing the faces will be easier for the next steps. Face landmark estimation is used to approach this process. There are a lot of ways to do this, one of the ways is using the method that was invented in 2014 by Vahid Kazemi and Josephine Sullivan [8]. The basic idea is the 68 points (facial landmarks) that exist on every face such as the top of the chin, the outside edge of the eyes, the inner edge of eyebrows, etc will be used and trained by machine learning algorithm.

2.3. Delaunay Triangulation and Warp Triangle

In the previous step, the facial landmark points are obtained from the face image. These points can be used to calculate a Delaunay triangulation. This model allows the image to be broken up into many parts of triangle. After the Delaunay triangulation model is calculated, the input face image is able to warped using this model.

2.4 Local Binary Pattern Face Recognition

The original LBP operator, introduced by Ojala et al. [9], is a powerful means of texture description. The operator labels the pixels of an image by thresholding the 3x3-neighbourhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor.

Later the operator was extended to use neighbourhoods of different sizes [10]. Using circular neighbourhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighbourhood. For neighbourhoods, we will use the notation (P, R) which means P sampling points on a circle of radius of R.

3. ALGORITHM

As mentioned in previous sections, pose and expression of the input image is a challenge in for recognition systems. Following up this consideration, this research has come up with a method which can solve this problem with a good accuracy in the result. The flowchart of this study is shown in Figure 1. In the first step, the images which have various poses and expressions are inputted into the system. The next step, the inputted image will be normalized for the expression and view pose by using the triangle transformation method. After the face pose view and expression are corrected, the mirroring method will be applied to correct the unnatural look of the output face image when the input face image has an extreme pose view angle. The neutral face reference is used as a template or reference for the neutral face. After these steps are passed, the performance of this system will be tested by comparing the modified face image with the virtual neutral face database. Local binary pattern (LBP) face recognition method is chosen for the method. The recognized image will be compared with the number of experiments to get the percentage of its performance.

3.1. Image Input with Various Poses and Expressions

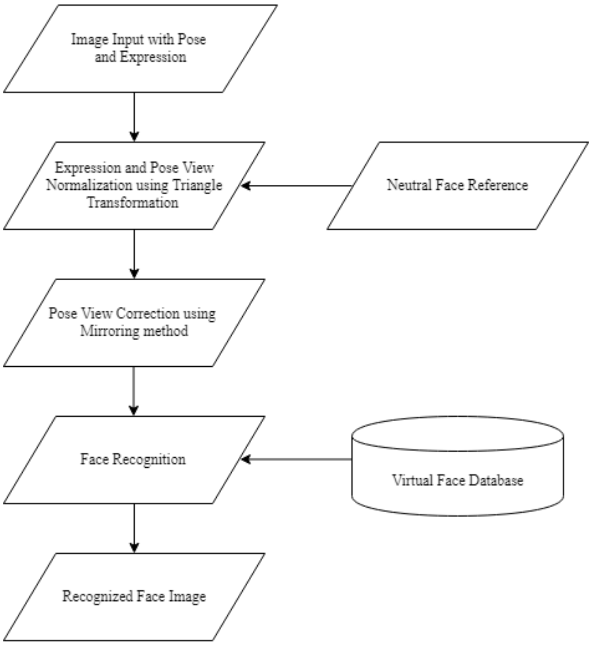

Figure 1. Overview of proposed method

There are two sets of images that are required for this research. Firstly, a set of face images that contains various expressions such as smile, scream, surprise, disgust, squint, etc. Not only expression but a variety of pose also needed to evaluate this system. To meet the requirements above, the Tarrlab face database is chosen. This face database was created at Brown University, USA. This database contains multiple images of 200 more individuals with many different races with consistent lighting, multiple views, real emotions, and disguises. For pose views this database has 5 different views, including 0 degree, +30 degree, -30 degree, +45 degree, -45 degree. Besides that, a virtual neutral face database is created for evaluating face recognition method with the result of the synthesized neutral face image. This gallery consists of the only neutral expression of the face image with a frontal view from the Tarrlab face database.

3.2. Expression Normalization Using Triangle Transformation

3.2.1 Detection

The very first step of this expression normalization procedure is detecting the facial features on the input image. The most common library for detecting face image; Dlib, is used for approaching this step.

Figure 2. The rectangle box shows the face that is identified as a face.

3.2.2 Facial Landmark and Cropping

The face image from the rectangle process before is used for this step. Facial points are placed in the face image to detect the eyes, nose, mouth, and shape of the face. Dlib provides 68 facial points. The Dlib library doesn't provide forehead points in their algorithm. However, these points are important for cropping the face image. So, five points are added manually by just using the top nose point and as a reference. These outer facial landmark points are used for cropping face image and will be applied into neutral face reference.

Figure 3. The face image with 68 facial landmark points and 5 additional points for the forehead.

3.2.3 Neutral Face Reference

One face image with the neutral expression and frontal view is needed to create the face reference model. This reference will be used a template for neutralizing the expression and pose of the input face image. First, the facial landmark points from the selected face image are obtained. Then, the points are transformed into a reference model by using Delaunay triangulation method. The Delaunay triangulation reference model is used as a single reference for all input images. All the input images will be transformed for the pose and expression following the shape of this Delaunay triangulation.

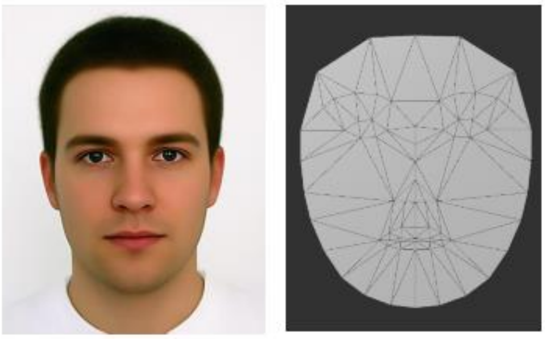

Figure 4. The face image that is used as a neutral face reference (left) and the Delaunay triangulation of neutral face reference (right).

3.2.4 Expression Normalization using Triangle Transformation

During this process, the facial landmarks of cropped input face image are projected onto the landmarks of Delaunay triangulation model. All the facial landmark points are synchronized. The input face image is warped to fit into the neutral face reference (Delaunay triangulation model). The eye and mouth expression of the input face image is transformed into neutral expression because of this process. As an illustration, if the mouth shape of the input face image is open, it will be closed because the mouth landmark points are projected following the mouth shape of the reference model which has a closed mouth shape.

This projected process also forces the pose view of the input face image to be in frontal face view. Then if the input face image has a pose that is not in frontal pose view, the result face image will be transformed into frontal pose view adapting the shape of Delaunay triangulation reference model.

Figure 5. The synthesized face image refers to the neutral face reference**.**

3.2.5 Pose View Correction using Mirroring Method

Regarding the previous step, the projected process forces the pose view of input face image to be in frontal face view following the shape of Delaunay triangulation reference model. This step shows a weakness in correcting pose view of input face image that has extreme pose view. For instance, when the pose view angle is more than 30 degree, extreme pose views from the input face image will affect to the quality of the output image. The blind side of the input face image that comes from the extreme pose views made the synthesized face image will look unnatural. Figure 6 illustrates how synthesized face image looks terrible.

In order to solve that problem, the mirroring method is used for the output face image. The idea of mirroring method is to take one side of a face from the output image. This half side then be copied and replacing the other side of output face image which looks unnatural. After the input face image is projected to single reference model, the system will detect the pose view degree of the output face image. The system is set, if the pose view angle from the output image is 30 degrees or more, the mirroring method will be activated. The example result of mirroring method can be seen in Figure 6.

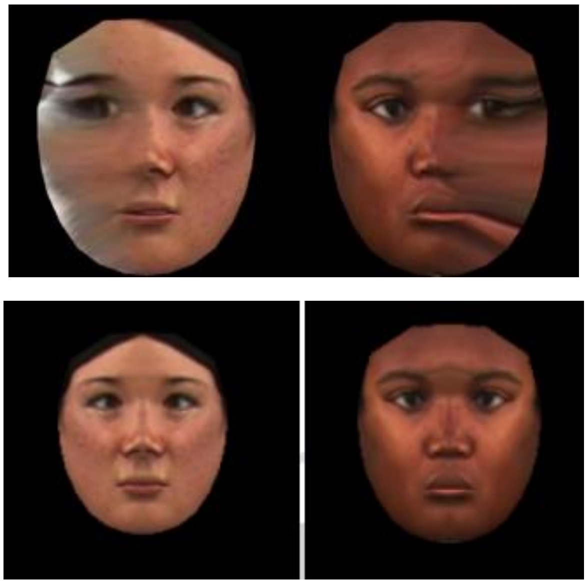

Figure 6. Comparison between the synthesized face image using the mirroring method and not using mirroring method**.**

3.2.6 *Face Recognition and Virtual Face Database*

The performance of this system is evaluated by using a face recognition scheme. The input face image that has been processed for the expression and pose needs to be compared to a set of original face images. The success of synthesized image recognizes the correct face image from database will be counted as a recognition rate.

Virtual face database is needed as a reference for face recognition scheme. This database contains a set of face images without expression and pose problems. All the images inside the virtual database have an only neutral expression and frontal face pose view.

4 ANALYSIS

4.1. Experimental Result

The experimental results of this method using the face database that contain pose views and expressions. The Tarrlab face database is used for this study. The face images are divided into two sets. First, sets of only neutral face image as a face database for the experiment. Second, sets of images that contain pose views and expressions. Every image in the database has four pose views (including expression for some images).

Table 1. Total images that is used for this research

Pose View Face Images and Expressions	**Face Image Database (Only Neutral Face)**	**Face Image Probe**
0 degree	241 images	241 images
+30 degree	-	241 images
-30 degree	-	241 images
+45 degree	-	241 images
-45 degree	-	241 images
		Total: 964 images

4.2. Face Recognition Performance

The result of this proposed method is tested by using a well-known algorithm method namely local binary pattern (LBP) method. Table 2 shows a recognition rate of face recognition accuracy using LBP algorithm. This proposed method is compared with the previous work [3] to see its performance. The previous study used the CMU Multi-PIE database for the face database.

Face recognition accuracy for proposed method has a higher result than previous work for all pose views. For 0 degree pose view, proposed method achieves similar rate with the previous work. On the other hand, the extreme pose view degree such as 30 and 45 degrees showed a lower accuracy than before (0 degree). Yet it still achieves a higher rate than the previous work. Therefore, the average face recognition of this proposed method also reaches a higher rate than the previous study.

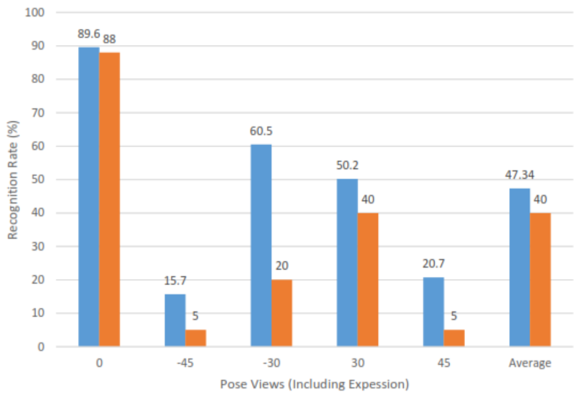

Figure 7. The Graph of Recognition Rate of Proposed Method using LBP

Table 2. The table of recognition rate of proposed method using LBP.

Pose View (Including Facial Expression)	**Recognition Rate (%)**	
	Proposed Method	**Previous Work**
0 degree	89.6	±88
+30 degree	15.7	>5
-30 degree	60.5	±20
+45 degree	50.2	±40
-45 degree	20.7	±5
Average	**47.34**	**±40**

4.3. Example of Neutralized Images

This proposed method warps the normalized pose view plus expression face to only neutral face reference to obtain a synthesized neutral face image for the frontal view with a neutral expression. This synthesized neutral image will provide a better similarity compare to virtual neutral face database. The example of this synthesized neutral faces that has front face pose view can be seen on Figure 8 and the synthesized neutral faces for the expression is shown on Figure 9.

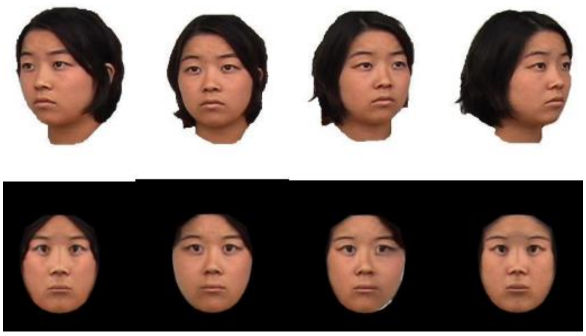

Figure 8. Example of Pose Views Correction Including -45, -30, +30, +45 degree View from Left to Right.

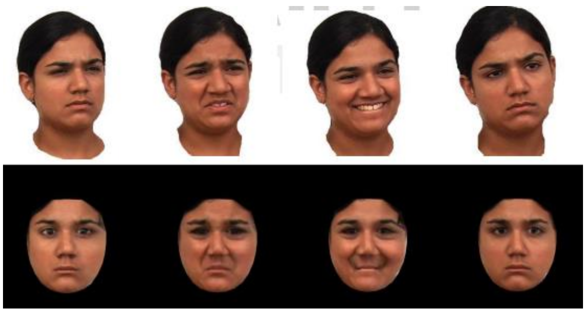

Figure 9. Example of Neutralized Expression Including angry, disgust, happy and sad from Left to Right

4.4. Conclusion

In this research, the simple method for solving problems of face recognition because of various pose and expression is proposed. To achieve this method, the input images are normalized for its pose and expression by using the triangle transformation. The single face reference is required as a template for transforming all input image.

Moreover, it is shown that this proposed method for transforming the face image with pose and expression into a neutral face can improve the face recognition accuracy rate by comparing it to the virtual neutral database. The improvement can be seen in face recognition accuracy rate using the well-known Local Binary Pattern (LBP) method. In all pose view including 0, -30, +30, -45 and +45 degrees, this proposed method has achieved a better result than the previous work [3] markedly for the extreme pose view like 30 and 45 degrees.

Yet, this proposed method needs to be reviewed in the future because the face database that is used in this research is different from the previous work. The Tarrlab face database is used for this research while CMU Multi-PIE database is used for previous work. The same face database can be used for the future work to get a fair comparison for the face recognition rate.

REFERENCES

[1] M. Ramachandran, S.K. Zhou, D. Jhalani, R. Chellappa, "A Method for Converting a Smiling Face to A Neutral Face with Application to Face Recognition", Proceedings of the IEEE Int. Conf. on Acoustics Speech and Sign. Proc., vol. 2, pp. ii/977-ii/980, May 2005.

[2] C.-K. Hsieh, S.-H. Lai, Y.-C. Chen, "Expression-Invariant Face Recognition with Constrained Optical Flow Warping", IEEE Transactions on Multimedia, vol. 11, no. 4, pp. 600-610, 2009.

[3] C. Petpairote, S. Madarasmi, K. Chamnongthai, “A Pose and Expression Face Recognition Method Using Transformation-Based On Single Face Neutral Reference”, Wireless Summit (GWS), 2017 Global, pp. 123-126, October 2017.

[4] Y. Hong, S.E. Choi, G.P. Nam, H. Choi, J. Cho, G.J. Kim, “Adaptive 3D Model-Based Facial Expression Synthesis and Pose Frontalization”, Sensors (Basel), 2020.

[5] B. Chu, S. Romdhani, and L. Chen. 3d-aided face recognition robust to expression and pose variations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1899–1906, 2014.

[6] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, Computer Vision and Pattern Recognition, 2001.

[7] N. Dalal, B. Triggs, “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 2005.

[8] V. Kazemi, J. Sullivan, “One Millisecond Face Alignment with an Ensemble of Regression Trees”, IEEE Conference on Computer Vision and Pattern Recognition, June 2014.

[9] Ojala, T., Pietik¨ainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29 (1996) 51–59

[10] Ojala, T., Pietik¨ainen, M., M¨aenp¨a¨a, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.

[11] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions and pose variations. IEEE transactions on pattern analysis and machine intelligence, 2013.

[12] T. Gass, L. Pishchulin, P. Dreuw, and H. Ney. Warp thatsmile on your face: optimal and smooth deformations for face recognition. In FG 2011).

[13] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen. Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities. In 18th IEEE ICIP, Sept. 2011.

[14] M. Yang, L. Zhang, J. Yang, and D. Zhang. Robust sparse coding for face recognition. In IEEE CVPR 2011.

[15] A. M. Martinez and R. Benavente. The AR face database Technical report, CVC, June 1998. [16] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In Automatic Face and Gesture Recognition, 2002.

[16] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. Image and Vision Computing, 2010.