Neural Network-based Face Pose Tracking for Interactive Face Recognition System

Balza Achmad[#], Kartika Firdausy[*]

[#] Department of Engineering Physics, Universitas Gadjah Mada, Jl Grafika 2, Yogyakarta, 55281, Indonesia
E-mail: balzach@ugm.ac.id

[*]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia
E-mail: kartika@ee.uad.ac.id

Abstract— Security system has long been a very important aspect in almost every field. Technology advancement has made biometric systems applicable to provide higher level of security. **Face recognition is one of the popular biometric** modalities. An interactive face recognition system has been developed to increase the yield of the face recognition yet still uses a simple recognition algorithm. The system provides responses to the user in order to obtain a frontal face image. A neural-based face pose estimation is proposed to generate the proper response based on the captured face image. The overall accuracy of the pose recognition is 95.6%.

Keywords— face landmark, face pose tracking, face recognition, neural network.

I. INTRODUCTION

We are now living in the information era, which means that information openness is very essential. However, information security is very important to ensure its availability, integrity and confidentiality [1]. Accessing information resources often need us to enter password or PIN, which are the weakest means of authorization systems. Higher level security systems include the use of parts of the user body or gestures, which may provide not only authorization but also authentication [2]. This so-called biometric systems possess one-to-one relationship between the users and the data that represents themselves. Many biometric modalities have been developed during the last decades, including finger print, face recognition, iris recognition, hand geometry, palm print, DNA, speech recognition and signature verification [1].

Among many modalities of biometric system, face recognition is one of the most popular authentication and authorization systems [3], despite some issues or problems that exist in it. These problems are arisen from internal and external factors that take place during the process. The internal factors comprise the color and texture of the face, face pose as well as facial expression. The external factors consist of the background scene as well as lighting environment [2-3].

Face color and texture usually can be considered as parts of the face that need to be recognized, assuming these factors do not change for each individual during a relatively long span of time. To reduce the effect of other factors, those factors are standardized. By setting constant background and lighting, the variation due to the external factors can be minimized. Facial expression can be reduced by telling the user to express normally during the authentication process. The pose can be standardized as well, although some researchers developed technique to extend the recognition technique to include face pose variations [4-7]. This paper, however, focuses on the uses frontal face pose as standard pose, hence simpler face recognition techniques are applicable.

Chai et al. found that frontal face pose gave a more accurate result compare to angled poses [4]. Table I shows that the bigger the yaw angle the less accurate the face recognition result. The best result is achieved when the face pose is frontal to the direction of the camera.

TABLE I
ACCURACY OF FACE RECOGNITION FOR VARIOUS YAW ANGLES [4]

Yaw angle (degree)	Accuracy (normalized to frontal)
-40	14%
-25	27%
-15	63%
0	100%
15	47%
25	26%
40	11.5%

In order to prevent the lessening of accuracy by the face yaw angle, our system provides a feedback mechanism to maintain the user face pose frontal during a recognition process. For that we need to estimate the face pose. Face pose estimation techniques have been developed by many researchers. Breitenstein et al. estimated the yaw, pitch and roll angle face pose based on nose shape [8]. Aghajanian et al. used non-overlapping grid of patches to model the face pose [9]. Zhang et al. developed a technique to estimate face pose using 2-D partitioned sparse representation classifier [10]. Firdausy et al. classified face pose based on face landmarks [2]. The criteria developed for this classification were determined heuristically, which in turn still gave some errors. In this paper we proposed the use of artificial neural network to classify the face pose.

Some researchers utilized neural network for face recognition and pose estimation. Huang et al [5], Latta et al. [11] and Nazeer et al. [12] used PCA to generate eigenface and combined with neural network for face recognition. Anam et al. used cropped face image as a feature and combined the neural network with genetic algorithm [13]. Asteriadis et al. convolved face images with a series of filters and fed it into a neural network for head pose estimation [14]. In this paper, we used landmarks coordinates for the neural network inputs.

II. Materials and Methods

Fig. 1 illustrates the schematic diagram of a common face recognition system [12]. The system has two main subsystems, i.e. face detection and face recognition. During face detection step, an image is captured by a camera. In this case, the image scene mostly includes the user's upper body. An algorithm is subsequently applied to check whether the captured image contains a face. Then the face part of the image is extracted. The size of the face image is usually arbitrary, depends on some factors such as the resolution of the camera as well as the distance between the camera and the user. Therefore, the face image needs to be normalized to a size that matches with the face image database. In the second step, a feature vector is extracted from the normalized face image. Afterward, by applying certain classification and comparing to the image face database, a face class is obtained.

In this paper, we add some steps in between the original two steps of the above face recognition system. Since we only want the frontal face image to be fed to the face recognition subsystem, we introduce pose estimation with a response to the system. After the face detection subsystem supplies normalized face image, the pose of the face is determined. If the pose is frontal, then the image will be inputted to the recognition subsystem. Otherwise, the system provides a response to the user in order to fix his/her pose until a frontal face image is gathered.
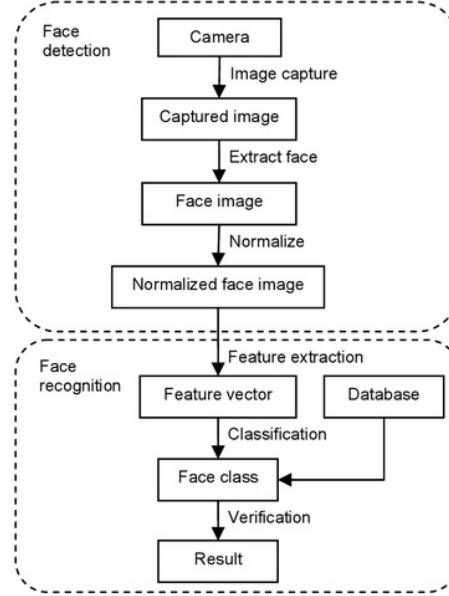

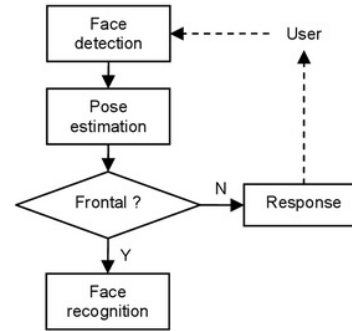Fig. 1 Face recognition system

Fig. 2 Interactive face recognition system with face pose tracking

The response is an action that depends on the pose captured by the camera, which is the opposite of the current pose. For instance, if the user face pose is turned to the left then the system will tell the user to turn his or her face to the right. If the user face pose is tilted to the right, the system will ask the user to tilt his or her face to the left. The face poses defined in this paper is given in Table II. Examples of the five face pose are given in Fig 3.

TABLE II
FACE POSES AND SYSTEM RESPONSES

Face pose	System response
Frontal	"Steady, please"
Tilted to the left	"Please tilt your head to the right"
Tilted to the right	"Please tilt your head to the left"
Turned to the left	"Please turn your head to the right"
Turned to the right	"Please turn your head to the left"

    (a) Frontal   (b) Tilted to the left   (c) Tilted to the right
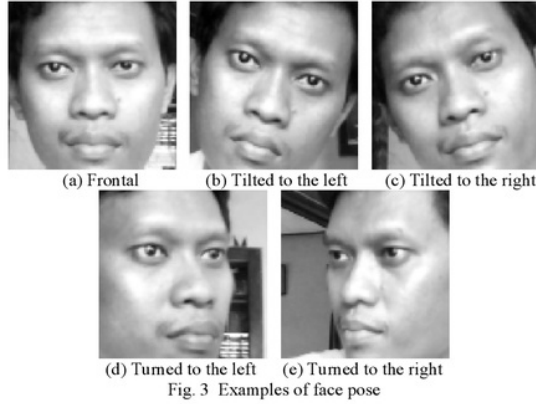
    (d) Turned to the left   (e) Turned to the right

Fig. 3 Examples of face pose

The proposed pose estimation subsystem firstly detects and calculates face landmarks (i.e. both eyes, nose and mouth) from face image (Fig. 4). The camera captures the scene. Once the user face has been detected, the face area of the image is cropped and for every detected face landmark, their coordinates are calculated. When all the four landmarks are detected, the result is then supplied to the neural network, otherwise, the step goes back to frame capture from the camera.
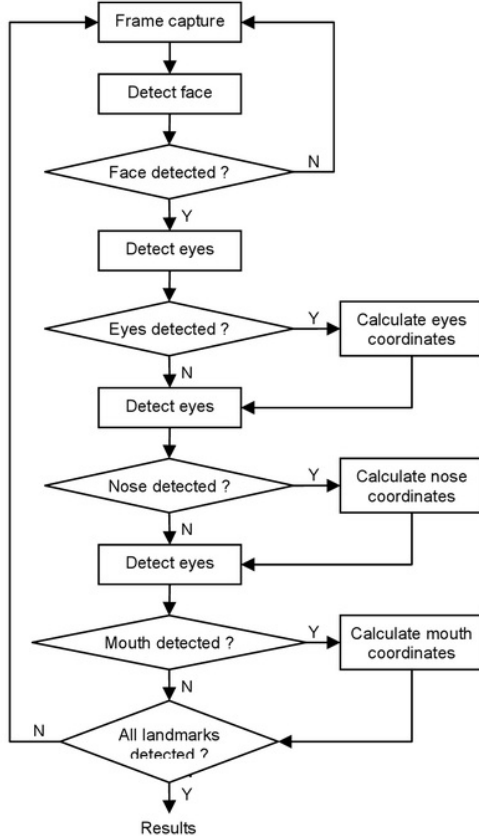

Fig. 4 Detection and calculation of face landmarks

The coordinates of the face landmarks are normalized to the width and height of the face image, hence the distance between the user and the camera will not produce a significant effect. The direction of the x coordinate is to the right, while the direction of the y axis is downward. The zero axis of the x coordinate is at the center of image face, whereas the zero axis of the y coordinate is at the top of the image.

The coordinates of the landmarks are then used for the input of the neural-based dace pose estimation (Fig. 2). Thus, the neural network has eight inputs, i.e. the x and y coordinates of four face landmarks. The neural network architecture is a feed forward network with one hidden layer consists of 10 nodes and logistic activation function. The output of the neural network is the face poses as has been defined in Table II.
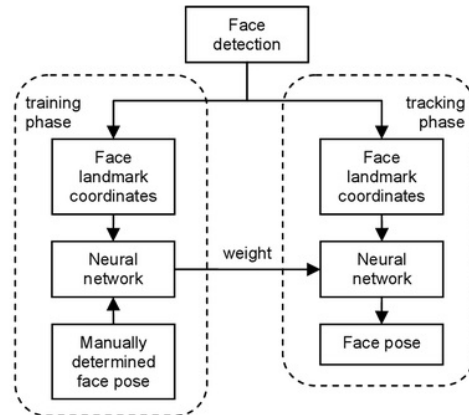

Fig. 5 The proposed neural-based face pose estimation

The output is a five-bit binary code for each face poses as described in Table III. The normalized face images from the face detection are divided into data for training and data for testing. During training phase, the system detects and calculates the face landmark coordinates, and the face poses are determined manually by the operator. Afterward, the neural network is undergone learning until convergence achieved. The weight of the trained neural network is then used for the tracking or operational phase in which the face poses will be calculated automatically by the neural network.

TABLE III
CODES FOR THE NEURAL NETWORK OUTPUT

Face pose	Ouput code
Frontal	10000
Tilted to the left	01000
Tilted to the right	00100
Turned to the left	00010
Turned to the right	00001

The user interface of the system is depictured in Fig. 6. It consists of the camera scene with detected face and landmarks, the cropped and normalized face image, the coordinates of the landmarks as well as the face pose. The system response is given in the form of speech voice.
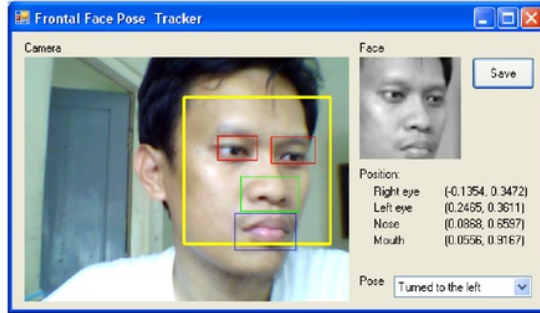
Fig. 6 System user interface

III. RESULTS AND DISCUSSION

We used 250 data for the training of which comprises 50 data for each pose. After the learning phase, another evenly distributed 250 data is used for testing. The result is given in Table III. The overall face pose estimation success rate is 95.6%.

TABLE III
CODES FOR THE NEURAL NETWORK OUTPUT

		Calculated pose				
		F	IL	IR	UL	UR
Actual pose	F	100%	-	-	-	-
	IL	4%	96%	-	-	-
	IR	6%	-	94%	-	-
	UL	2%	-	3%	95%	-
	UR	3%	4%	-	-	93%

Notes: F = Frontal, I = Tilted, U = Turned, L = Left, R = Right

It can be seen that the neural network can recognize completely the frontal pose. Whereas for other poses, some incorrect recognitions exist. Some of the non frontal poses are mistakenly identified as frontal pose. Some turned poses are also incorrectly recognized as tilted because there are some turned poses that nearly close to the corresponding tilted poses, i.e. turned to the left is similar to tilted to the right and vice versa.

This paper takes into account the face yaw, pitch and roll angles altogether. The turned poses are the effect of yaw and pitch angles, while the tilted poses are caused by the yaw and roll angles. However, there are overlaps between the tilted and turned poses on some yaw-pitch and yaw-roll angle combinations. Hence, the result depends on the manual face pose decision during the neural network training.

IV. CONCLUSIONS

A neural network based face pose estimation has been developed to provide responses in an interactive face recognition system. The face poses defined in this paper are frontal, tilted and turned for both directions to the left and to the right. The feed forward type neural network uses eight face landmark coordinates for the inputs and five face pose codes for the outputs. The success rate of the face pose estimation is 95.6%.

REFERENCES

[1] D. Bhattacharyya, R. Ranjan, F.A. Alisherov, M. Choi, "Biometric authentication: a review", *International Journal of u- and e- Service, Science and Technology*, vol. 2, no. 3, pp. 13-28, Sept. 2009.

[2] K. Firdausy, B. Achmad, "Automatic frontal face pose tracking for face recognition system", *International Journal on Advanced Science, Engineering and Information Technology*, vol. 1, no. 4, pp. 399-402, 2011.

[3] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "Face recognition: a literature survey", *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, Dec. 2003.

[4] X. Chai, S. Shan, S. W. Gao, "Pose normalization for robust face recognition based on statistical affine transformation", in *Proc. of the ICICS-PCM 2003*, paper 3A4.1, 2003.

[5] F. J. Huang, Z. Zhou, H. J. Zhang, T. Chen, "Pose Invariant Face Recognition", in: *Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.245-250, 2000.

[6] X. Chai, S. Shan, X. Chen, W. Gao, "Locally linear regression for pose-invariant face recognition", IEEE Trans. on Image Processing, vol. 16, no. 7, pp. 1716-1725, Jul. 2007.

[7] X. Zhang, Y. Gao, "Face recognition across pose: a review", *Pattern Recognition*, vol. 42, pp. 2876-2896, 2009.

[8] M. D. Breitenstein, D. Kuettel, T. Weise, L. van Gool, H. Pfister, "Real-time face pose estimation from single range images", in *Proc. CVPR 2008*, pp. 1-8, 2008.

[9] J. Aghajanian, S. J. D. Prince, "Face pose estimation in uncontrolled environments", *in Proc. of the British Machine Vision Conference (BMVC) 2009*.

[10] C. Zhang, Y. Zhang, L. Liao, "Two dimensional partitioned sparse representation for head pose estimation", in *APSIPA Annual Summit and Conference (APSIPA ASC)*, 2011.

[11] P. Latta, L. Ganesan, S. Annadurai, "Face recognition using neural networks", *Signal Processing An International Journal (SPIJ)*, vol 3, issue 5, pp. 153-160, Nov. 2009.

[12] S. A. Nazeer, M. Khalid, *PCA-ANN face recognition system based on photometric normalization techniques*, in State of the Art in Face Recognition: I-Tech, 2009.

[13] S. Anam, M. S. Islam, M. A. Kashem, M. N. Islam, M. R. Islam, M. S. Islam, "Face recognition using genetic algorithm and back propagation neural network", in Proc. of the International MultiConference of Engineers and Computer Scientists 2009 Vol I, Mar. 2009.

[14] S. Asteriadis, K. Karpouzis, S. Kollias, "Face tracking and head pose estimation using convolutional neural networks", in *Proc. of the ACM/SSPNET 2nd International Symposium on Facial Analysis and Animation*, 2010.

Ijaseit_164_175_1_PB.pdf

9 Kartika Firdausy, KZ Widhia Oktoeberza. "Segmentation of optic disc using dispersive phase stretch transform", 2016 6th International Annual Engineering Seminar (InAES), 2016
Crossref

10 words — < 1%

10 www.islab.ntua.gr
Internet

8 words — < 1%

11 www.wseas.us
Internet

8 words — < 1%

12 oda.hio.no
Internet

8 words — < 1%

13 www.umiacs.umd.edu
Internet

8 words — < 1%

14 liris.cnrs.fr
Internet

8 words — < 1%

15 publications.im2.ch
Internet

8 words — < 1%

16 Hui Qi, Qingjie Zhao, Xiongpeng Wang, Mingtao Pei. "Pose-indexed based multi-view method for face alignment", 2016 IEEE International Conference on Image Processing (ICIP), 2016
Crossref

7 words — < 1%

17 Thiemo Wiedemeyer. "Wide Range Face Pose Estimation by Modelling the 3D Arrangement of Robustly Detectable Sub-parts", Lecture Notes in Computer Science, 2011
Crossref

6 words — < 1%