ISBN 978-93-86878-07-6
10th International Conference on Biological, Biomedical, Medical and Pharmaceutical Sciences (Cellular and Molecular Approach) (BBMPS-17)
Dubai (UAE) Dec. 21-22, 2017

Gaze Tracking Algorithm using Neural Network

Bich Lien Nguyen[1] and Van Anh Nguyen[1]

[1]Hanoi University of Science and Technology, Hanoi, Vietnam

Abstract**:** *In this paper, an efficient methodology of tracking eye gaze by using neural network is proposed. Using eye features which relate to the movement of the eyes as inputs, a neural network is built to recognize the gazing point that a participant is looking at on the screen. With the final results of 91.25% sensitivity and 81.97% specificity on the data set from 6 participants, it is established that the proposed method can successfully track eye gazing.*

Keywords: *Gaze Tracking, Neural Network, Feature Extraction*

1. Introduction

Gaze tracking is normally defined as the process of estimating and locating the gazing point or the point that a person is looking at. For many people with disabilities and the elderly who are not able to talk or use their hands, a gaze tracking device which can help them communicate via a human-computer interface may become an essential part of life.

A variety of approaches have been introduced to accomplish the task of gaze tracking. Some of them require attachments to the eyes or the head [1,2] while others rely on images of the eye taken by cameras without any physical contact. It is certain that wearing instrusive devices like contact lenses, electrodes or headgear is uncomfortable, especially for disabled people. Camera-based techniques efficiently help to overcome that inconvenience, thus has recently become the favourite solution in the area of gaze tracking. A common method to implement camera-based gaze tracking is applying infrared light sources to users' eyes and using special cameras to capture reflected images of eyes [3,4]. By this way, effects of surrounding light conditions can be eliminated and better image quality can be obtained, so that the accuracy of gaze tracking can be enhanced. However, infrared lights, in sufficient concentrations, not only cause inconvenience but alo damage to human eyes. Besides, the high cost of an infrared system can be another disadvantage of this type of system.

The main objective of this paper is to propose a method of gaze tracking using images captured by a common digital camera without requiring any attachment or infrared equipment. The paper is divided into four sections. Section 2 provides an overview of the methodology used for gaze tracking using neural network. In Section 3, the development and results of the study will be mentioned and discussed. Conclusions for this study are drawn in Section 4.

2. Method

2.1. Experiment

There are 6 people participating in our experiment. A digital camera is mounted on the top of a screen to continuously capture images of participants. At this stage of our study, the screen of a 24-inch computer monitor is utilized. This screen is designed as a keypad with 6 different keys which are labelled as letters A, B, C, D, E, F as shown in Fig.1. Each participant is asked to implement the following procedures: (i) continuously gaze at each key on the screen for 30 seconds, (ii) look at each key for 10 seconds and consequently move to all other keys, keeping at each for 10 seconds. After different trials, the distance from the sitting position of participants

to the camera is set at 0.8 meter which is determined to provide images with acceptable quality without making any inconvenience to the participants.

The participants are asked to try their best to keep their head position unchanged to the screen and look at the center of each key when doing experiment procedures. During the experiment, neutral lighting and quite ambience are maintained so that participants can concentrate on doing tasks.

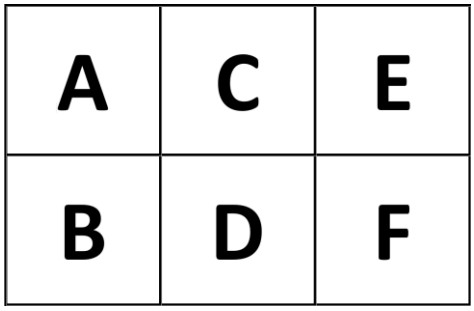

Fig. 1: Design of screen as a keypad with 6 different keys

2.2. Proposed Algorithm for Gaze Tracking

The proposed method for tracking eye gaze is illustrated in Fig.2. First, using data obtained by the camera, an algorithm is applied to detect the face and locate the position of the eyes. From the processed eye images, eye features which relate to the movement of eyes will be extracted. Using these features as inputs, a neural network is built to recognize the gazing point that the user is looking at on the screen.

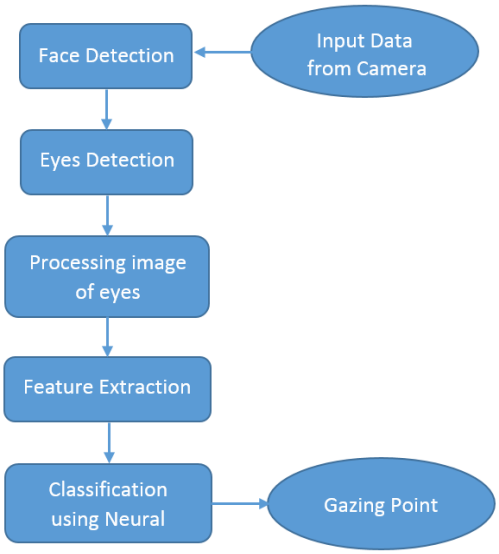

Fig. 2: Process of the proposed gaze tracking algorithm

2.2.1. Face and Eyes Detection

Using the images captured from camera, the face and eyes detection step is implemented. The aim of this step is locating first the position of the face and then the positions of the two eyes on a participant's face. To do this, we apply Viola-Jones algorithm which is well known as the most popular and efficient algorithm for recognizing human face [5,6]. In this algorithm, instead of using raw pixel values, Haar-like features are estimated from intergral image which converted from gray-scale image from camera. These features are used as inputs to a cascade of classifiers built based on AdaBoost Learning Algorithm which selects the most efficient group of features for detecting objects. The cascade structure combines classifiers in a way that significantly boost the speed of the detector by focusing in promising face-like and eye-like regions of the image.

2.2.2. Feature Extraction from Eyes

After detecting regions of the left and right eyes, the step of extracting features are implemented. In this study, eye features are extracted based on the eye deformable template which describes geometric characteristics of human eyes by characterizing a set of 11 parameters which are able to vary during the changing of the eyes' state and position [7]. In this template, the bounding contour of the eye is modelled by two parabolic curves while the boundary between the iris and the white of the eye is modelled by a circle. Then a set of 11 features, represented by $g = (x_c, y_c, x_e, y_e, p_1, p_2, r, a, b, c, \theta)$, is defined as in Fig. 3.

As a result, a total set of 22 eye features (11 different kinds of feature x 2 eyes) are estimated for each image. To reduce the variability in data, each final data point is estimated as the average of two consecutive non-overlapping points. Statistics is then applied to determine the significance of changes of these features when the two eyes move to 6 different positions of the screen designed in part 2.1. In all statistic tests, probability values (p-values) less than 0.05 are considered to be significant.

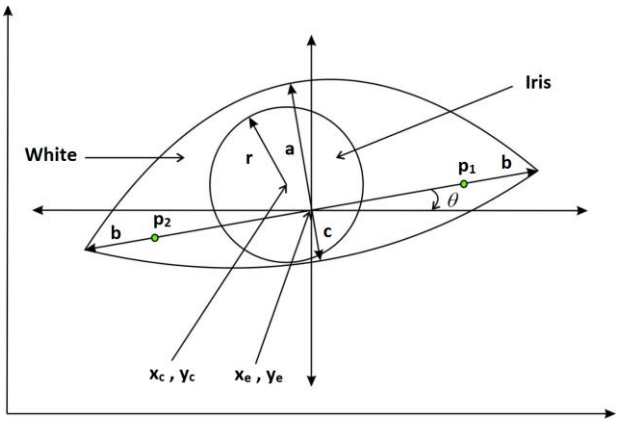

Fig. 3: Deformable template for a human eye [7]

2.2.3. Classification using Neural Network

Artificial neural networks have been employed popularly in biomedical area as a powerful tool of classification and pattern recognition which can effectively model non-linear relationships between inputs and outputs.

In this study, a neural network with feed-forward multilayer structure is developed as a classification unit. The network consists of one input layer, one hidden layer and one output layer. The input layer includes 22 nodes which are the 22 eye features extracted in part 2.2.1. The output layer includes 6 nodes corresponding to the states of 6 gazing points on the screen. Each output is set at 1 if the participant is gazing at the corresponding point on the screen and -1 if the participant is looking at the other points.

The overall data of each participant are grouped into a training set, a validation set and a testing set. The developed neural network is trained by using the training set with a stopping procedure determined by the validation set. This neural network is trained by the Levenberg-Marquardt (LM) algorithm which is an effective and popular training algorithm. In brief, the LM algorithm estimates the second directional derivative of the error function, in order to direct the training process to a local minimum and find optimized network parameters. The number of hidden nodes is selected as the one which give the best classification performances.

After determining the final structure and parameters for the neural network, the Receiver Operating Characteristic (ROC) curve will be found based on the combined training and validation dataset. By definition, a ROC curve presents the tradeoff between the true positive rate versus false positive rate (equivalently, sensitivity versus 1–specificity) for different thresholds of the classifier output. In this paper, we use that characteristic to choose the output threshold for defining the two states of gazing and non-gazing for each output.

Finally, based on the derived structure as well as the output threshold of neural network, the test set will be applied to test the performance of the developed neural network. The performance of each output is determined in terms of sensitivity and specificity of the classification results:

$$Sensitivity = \frac{TP}{TP + FN} \quad ; \quad Specificity = \frac{TN}{TN + FP}$$

where

- True Positive (TP) is the number of gazing points which are correctly classified as gazing
- True Negative (TN) is the number of non-gazing points which are correctly classified as non-gazing
- False Positive (FP) is the number of gazing points which are wrongly classified as non-gazing;
- False Negative (FN) is the number of non-gazing points which are wrongly classified as gazing.

3. Results and Discussions

Statistical results show significant changes in extracted eye features when the eyes of participants move to different keys designed on the screen. Consistently with both left and right eyes in all 6 participants, the center of iris is established to be the most significantly features ($p<0.0001$). Slight changes ($p<0.01$) are found in features that model the bounding contour and the whites of the eye. These results are quite similar when comparing data between the left eye and the right eye of each participant. However, these changes are different between participants (significant in some participants and not significant in others). It should be noted that the achieved statistical results are predictable because the eyes of each person has different shape and characteristics. Based on these results, we establish that extracted eye features are related to the eye movement and can be utilized to track eye gazing.

A neural network is developed with 22 input nodes (11 features x 2 eyes), 6 output nodes (6 gazing points on the target screen) and S hidden nodes. In this study, S is varied from 6 to 12 to select the one the give the best classification performance. As a result, the final network structure with S = 10 is determined as the one that yields the best result. The overall data of each participant is separated into a training set, a validation set and a testing set. The training set and validation set is acquired when the participants implement the experimental procedure (i) while the testing set is acquired when they implement the experimental procedure (ii) (refer to section 2.1 about experimental procedures). After the network is trained and validated by the data from the training set and validation set of each participant, all testing classification results corresponding with this network structure are reported in Table I.

The results for each output of the network (corresponding to each gazing position on the screen) show that classification performances are quite consistent between all six participants. For all participants, results of keys A, B (keys on the left of the screen) and keys E, F (keys on the right of the screen) are slightly higher than results of keys C, D (keys at the middle of the screen). These results are established to be suitable with the fact that when the eyes gaze to corners of the screen, the eye features will have more significant changes, thus it will be easier to track than when the eyes gaze to the center positions.

TABLE I: Classification Results

	Key A		Key B		Key C		Key D		Key E		Key F		Average Results	
	Sen	Spe	Sen	Spe	Sen	Spe	Sen	Spe	Sen	Spe	Sen	Spe	Sen	Spe
Participant 1	90	79	92	82	86	78	89	81	93	83	92	82	90.33	80.83
Participant 2	94	83	96	86	89	80	92	79	92	85	96	85	93.17	83.00
Participant 3	94	88	92	87	90	80	92	80	94	89	90	88	92.00	85.33
Participant 4	88	78	90	78	82	75	84	78	92	73	92	78	88.00	76.67
Participant 5	96	86	96	88	90	84	93	82	95	82	95	86	94.17	84.67
Participant 6	90	82	93	82	84	80	88	82	92	82	92	80	89.83	81.33
Final results of the study													**91.25**	**81.97**

Sen: Sensitivity; Spe: Specificity

There are also differences between the results of six participants. It is shown in Table I that slightly better results are achieved for participants 2, 3, 5 compared to the results of participants 1, 6. While the best results are achieved with participant 5 (94.17% sensitivity and 85.67% specificity), the worst results are achieved with participant 4 (88% sensitivity and 76.67% specificity). The mentioned trends are consistent for all tasks of gazing at six keys on the screen. It should be noted that these results are understandable because characteristics of the eyes are unique and different from person to person, so that it is impossible to achieve the same results for all people. It is expected that a more advanced classification algorithm will help to overcome this and enhance the overall results of the proposed algorithm.

4. Conclusions

In this paper, a method for eye gaze tracking using neural network is proposed. Statistical results show that extracted eye features are significantly related to the movement of the eyes. With the average classification results of 91.25% sensitivity and 81.97% specificity on the testing set, it is shown that the developed neural network-based classification unit is efficient in tracking eye gaze. Based on the developed neural network, in future work, more advanced techniques to train and optimise network structure can be explored in order to improve the overall accuracy of the system.

5. Acknowledgements

This research is funded by Hanoi University of Science and Technology (HUST) under grant number T2016-PC-092.

6. References

[1] Y. J. Ko, E. C. Lee, and K. R. Park, “A robust gaze detection method by compensating for facial movements based on corneal specularities,” *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1474-1485, July 2008.

[2] K. Takemura, K. Takahashi, J. Takamatsu, and T. Ogasawara, “Estimating 3-D point-of-regard in a real environment using a head-mounted eye-tracking system,” in *IEEE Trans. Human Mach. Syst.*, vol. 44, no. 4, pp. 531–536, August 2014.

[3] A. Haro, I. Essa, and M. Flickner, “A non-invasive computer vision system for reliable eye tracking,” in *Proc. Extended Abstracts Human Factors Comput. Syst.*, 2000, pp. 167–168.

[4] H. C. Lee, D. T. Luong, C. W. Cho, E. C. Lee and K. R. Park, "Gaze tracking system at a distance for controlling IPTV," in *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2577-2583, November 2010.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, pp. I-511-I-518 vol.1.

[6] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," *Proceedings. International Conference on Image Processing*, 2002, pp. I-900-I-903 vol.1.

[7] A. L. Yuille, D. S. Cohen and P. W. Hallinan, "Feature extraction from faces using deformable templates," *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on*, San Diego, CA, 1989, pp. 104-109.