Advanced Science and Technology Letters
Vol.86 (Ubiquitous Science and Engineering 2015), pp.85-90
http://dx.doi.org/10.14257/astl.2015.86.18

Gaze Recognition Based on Correlation of Facial Components

Ji-Su Yang[1], Kwang-Seok Hong[1]

[1] College of Information & Communication Engineering,
Sungkyunkwan University Suwon, 440-746, South Korea
yamido@skku.edu, kshong@skku.ac.kr

Abstract. In this paper, we propose a fast and accurate gaze recognition method by using the correlation of a small amount of computation facial components. The proposed method detects the face, eyes, nose and mouth. In the detected face region, using the Haar-like features, the center point of the detected area, the distance between each of the components, and each of the components was extracted through the angle of the three corners of the triangle in the components of the detected face. The extracted features in the learning set were used for recognition experiments with the Random Forest algorithm, achieving a very high recognition rate of 99.46% for the recorded results. In addition, the perform rate was demonstrated to be 23.6fps.

Keywords: Gaze estimation, Face detection, Facial component, Haar-like feature, Random Forest

1 Introduction

Research on gaze recognition technology is actively being conducted for application to a variety of fields, including Human Computer Interaction (HCI), advertising strategies, intentions, etc. HCI is the study of how people can interact with computer systems in an easy and convenient fashion, and the design of related systems, in which evaluation of human characteristics can be considered a part. Further, gaze recognition technology is also being developed due to widening of the applications, giving it increased attention due to application to various fields for accurate eye detection [1].

2 Related work

In the existing recognition research of Vatahska, Bennewitz, and Behnke [2], a method using Haar-like feature and unique facial feature (nose, eyes, ears) detection was proposed, based on neural network recognition through the gaze direction. In addition, Chen and Lemon [3] detected the face and eyes using Haar-like features of the Cascade method. To date, 3000 eye and 8000 non-eye images, 18 * 12 in size,

ISSN: 2287-1233 ASTL

were used for study. An entropy analysis method using a histogram was used to detect the corners of the eye in the detection region, while the ends of the mouth also utilized similar methods. Nostrils were detected through assumption of the dark areas in the nose. The feature points detected are tracked using the Optical Flow (LK), the head pose is then estimated by POSIT [4]. However, the above methods have the disadvantage that a large amount of computations need to be performed in real time, or slow processing due the requirement for real 3D matching points. To overcome these disadvantages, gaze direction was recognized herein using the correlation of the components of the face, which allowed reduction of the processing time of the algorithm and increase of the recognition accuracy.

3 Proposed method

3.1 Face and Facial components Detection

In this paper, the face and facial components (eyes, nose, mouth) were detected using Haar-like features [7], and a learning classifier through the Adaboost algorithm.

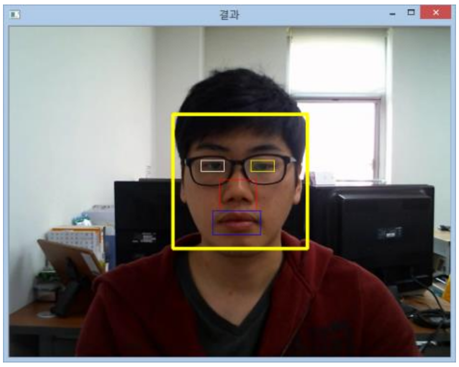

Fig. 1. Face and facial component detection results

Fig. 1 shows a facial image obtained via webcam, and the facial components of the eyes, nose and mouth detected in the image. The yellow box indicates the face region detected through the learned classifier in the Adaboost algorithm, which extracted the detected regions to specify the region of interest (ROI). The facial components were then extracted from the face images. The face typically consists of two eyes, a nose, and a mouth, positioned mostly the same. The face image was split into the eyes, nose and mouth according to the position to increase the detection accuracy and speed using this feature, as follows:

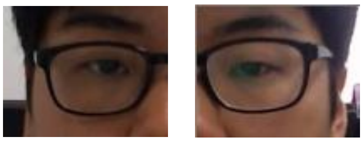

Fig. 2. Split image for Eye Detection

Fig. 3. Split image for Nose Detection

Fig. 4. Split image for Mouth Detection

Fig. 2 illustrates the split images of the eye position, done in order to reduce the error detected on one side by sensing the image segmentation of the eye for eye detection, doubled to the left or right. Fig. 3 and 4 show the images for mouth and nose detection, for which a learned classifier in Adaboost was implemented.

3.2 Feature Extraction

In this paper, a trained classifier via Haar-like features in Adaboost was used to extract 32 features in the component faces. The face components were detected, and the features were extracted as follows:

a. Each component (left eye, right eye, nose, and mouth) and the distance between the face center point.

b. The distance between each component (left eye, right eye, nose, mouth).

c. Each component of the (left eye, right eye, nose, mouth), forming a triangle with three angles.

3.3 Gaze Direction Recognition Method

In this paper, recognition experiments were performed using the Random forest [8] algorithm, analyzing facial images taken with a camera installed on top of a monitor in 9 directions of rotation, including upper left / front / top / upper right / left / front /

right / bottom left / bottom / left-bottom. Video was shot for one minute each for the learning and recognition experiments.

4 Experiments and Results

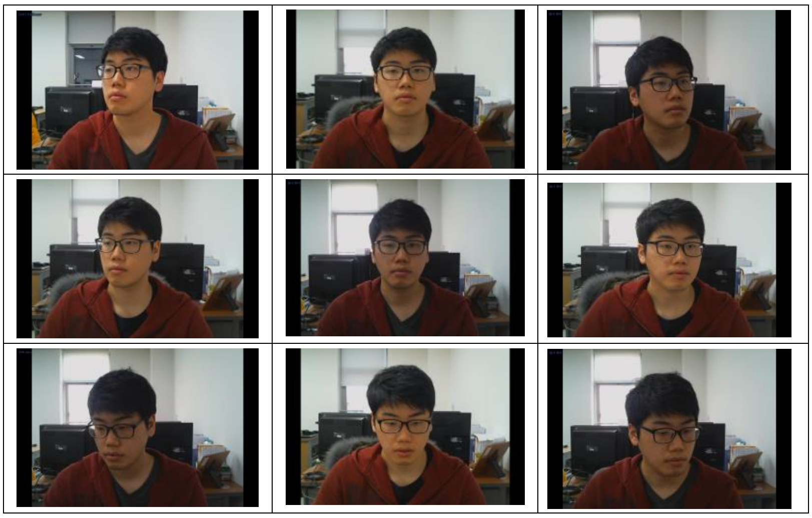

Fig. 5. Images captured in 9 directions video

Fig. 5 shows the images of nine directions used in this study. The detection counts of each image for the face and facial components were as follows.

Table. 1. in each direction face and facial component detection times

	Upper left	Top	Upper Right	Left	Front	Right	Bottom left	Bottom	Bottom Right
Frame	595	687	549	582	577	573	602	584	581
Face	375	687	393	240	577	554	600	584	484
Left eye	369	636	375	230	564	512	422	561	326
Right eye	370	476	377	230	567	491	581	556	195
Nose	372	684	393	228	576	553	600	584	484
Mouth	364	687	377	169	576	533	481	526	468

Using different combinations of seven features, the performance of the experiment was measured from the above features. The following results were obtained:

a. Eyes, nose and mouth of each of the center points / between distance eyes, nose and mouth / eyes, nose, mouth, each consisting of a triangle (total of 32)
b. Eyes, nose, mouth, each consisting of a triangle (total of 18)
c. Eyes, nose, mouth, each consisting of a triangle (total of 12)
d. Distances between the eyes, nose and mouth / eyes, nose, and mouth, each consisting of a triangle (total of 24)
e. Distances between the eyes, nose and mouth / eyes, nose, and mouth, each consisting of a triangle (total of 12)
f. Distances between the eyes, nose and mouth / eyes, nose, and mouth, each consisting of a triangle (total of 12)
g. Eyes, nose, and mouth, each consisting of a triangle (6 total)

Table. 2. Nine gaze direction recognition experiments using Random forest

	a	b	c	d	e	f	g
100:100	99.98%	99.83%	99.88%	99.91%	99.72%	99.79%	97.91%
7:3	99.46%	98.68%	98.68%	98.58%	98.14%	98.53%	88.86%
Frame rate	23.6fps (42ms)						

The learning and testing data rate 100:100 in more than 97% during the experiment, 7:3 was recorded during the recognition rate over 88% in experiment, the highest recognition rate achieved was 99.46%. In addition, the speed is suitable for real-time processing, at 23.6 fps.

5 Conclusion

In this paper, the facial components were detected from images obtained through a web camera using the proposed method for recognizing the gaze direction. The proposed method employed Haar-like features for detection of the face and facial components (both eyes, nose, mouth), the center point in the detected component, the distance between each of the components, and the components that make up each triangle with the extracted features of each. The extracted features were used in the learning and recognition experiments with the random forest algorithm. From 9 recognition results, an accuracy of 99.46% was obtained. Also, the performance rate demonstrated suitable speed for real-time processing, at 23.6fps. Future studies should be carried out with consideration of a number of different environments in order to improve the disadvantages due to limited position and distance of the camera. In addition, face research is needed to recognize the direction of the user's gaze.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2014023704)

References

1. Csikszentmihalyi, Csikszentmihalyi.: Optimal experience. UK. Cambridge University Press (1988)
2. Teodora, V., Maren, B., Sven, B.,: Feature-based Head Pose Estimation from Images. 7th IEEE-RAS International Conference on, pp 330--335 (2007)
3. Qian, C., Haiyuan, W., Takeshi, F., Masahiko, Y.,: 3D Head Pose Estimation without Feature Tracking, Automatic Face and Gesture Recognition, Proceedings. Third IEEE International Conference on, pp88-93 (1998)
4. Wan-zhi, Z., Zeng-cai, W., Jun-kai, X., Xiao-yan C.,: A Method of Gaze Direction Estimation Considering Head Posture. International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.6 No.2, pp.103--112 (2013)
5. Jingying, C., Oliver, L.,: Facial Feature Detection and Tracking in a New Multimodal Technology-Enhanced Learning Environment for Social Communication. IEEE International Conference on Signal and Image Processing Applications. (2009)
6. Dementhon, D., Davis, L.,: Model-Based Object Pose in 25 Lines of Code. International Journal of Computer Vision, volume 15, pp 123--141 (1995)
7. Viola, P., Jones, M.J.,: Robust real-time object detection. Technical Repart Series, Comp-aq Cambridge research Laboratory (2001)
8. Breiman, L.,: Random Forests, Machine Learning. Volume 45, Issue 1, pp 5--32 (2001)

