Eye Shape Detection Based on Texture Modeling in Log-Polar Domain Using Principal Component Analysis and Recurrent Neural Network

MOHAMMAD ALDIBAJA^{*†} Member, SHINICHI SUZUKI[†] Non-member

(Received January 5, 2015, revised February 12, 2015)

Abstract: This paper presents a method to detect the eye contour using recurrent neural network. Log-Polar transform is used to convert the changes of eye rotations and scales into translations. Principal component analysis is applied on the eye RGB color channels in order to obtain a stable texture representation regardless lighting conditions. A recurrent neural network is trained to discern the patterns of texture values around the eye contour as well as to model the eye contour appearance at different eye states. The experimental results have verified that the recurrent network is able to detect the eye contour robustly. The proposed method outperforms the conventional methods in terms of accuracy and detection time.

Keywords: Eye Feature Extraction, Object Shape Detection, Recurrent Neural Network, Log-Polar Transform, Principal Component Analysis

1. Introduction

Human-Machine Interface (HCI) systems have become one of the most attractive fields in image processing [1] [2]. Eye is a very promising interactive component for implementing such a system. It has been proven that using eyes in interface systems is faster than hands [3]. Eye features such as eye corners, gaze point, pupil location, eye contour and blinking frequency can be employed in many applications. Nowadays, car companies try to increase the human safety by tracking the blinking frequency in order to detect the driver fatigue [4]. The handicapped people who have problems in controlling wheel chair by hands can be helped using eye gaze estimation techniques [5].

Eye contour is considered as the key-point to extract the eye features. Many methods have been proposed for feature extraction in the literature [6]~[9].

Active shape model (ASM) is one of the most popular techniques that can be used for eye contour detection [6]. Active Appearance Model (AAM) has been presented as an advanced version of ASM by modeling the object's shape and texture [7]. Yuile and et al. have proposed Deformable Template (DT) to detect the eye contour using an adaptive template [8].

Most of the above methods suffer from the difficulties of detecting the eye contour under different rotation angles, scales, eye states, lighting conditions and translations. Moreover, the essence of these algorithms is to search on the eye contour iteratively. The number of iterations is not fixed and depends on the complexity degree of the eye structure. Therefore, detection time is not fixed and these meth-

ods should be improved for using in real time applications such as interactive systems.

In this paper, the proposed method is inspired by our previous work [10]. The eye contour is represented by a set of points that form the eye shape. The distribution of these points represents the lower and upper eyelids as well as the two eye corners. The eye image is transformed into Log-polar domain in order to make the proposed method scale-rotation invariant [11]. A stable representation of the eye contour against wide range of lighting conditions is guaranteed using Principal Component Analysis (PCA). PCA is applied to highlight the variance between eye RGB color channels using the first eigenvector [12]. Each row in the first eigenvector image is called pixel profile and can be indicated by the corresponding θ coordinate. Each pixel profile represents the eye contour by a few pixels. A recurrent neural network is trained to model and analyze the patterns of these pixels. The trained network tries to predict the location of the eye contour in pixel profiles using some previous/past input-output status.

The experimental results have verified a better performance of the proposed method compared to ASM and DT in terms of accuracy, stability and detection time.

The reminder of this paper is organized as follows: In section 2, a brief explanation on the proposed solution to treat the eye contour under different scales, rotation angles and lighting conditions is provided. Section 3 describes the structure of the recurrent neural network in order to model the pixel values in the Log-Polar eye images. Setup and experimental results are declared in section 4. The obtained results are discussed in section 5 and the essence of the proposed method with highlighting the yielded advantages are concluded in section 6.

* Corresponding: maddib@yahoo.com

[†] Department of Mechanical Engineering, Toyohashi University of Technology, Toyohashi 441-8580, Japan

2. Eye appearance in Log-Polar domain

2.1 Log-Polar transform Log-Polar Transform (LPT) has been employed in many image processing applications such as optical flow [13], active vision systems [14], pattern recognition [15],...etc. An image in Cartesian coordinate system is exponentially sampled around the image center. Meanwhile, the eye image is radially divided into sectors. Based on the sector's radius and angle, a pixel $I = (x, y)^T$ is transformed into a pixel $\hat{I} = (r, \theta)^T$ in the Log-polar image using Eq. (1). The width of the Log-Polar image determines the ring sampling rate whereas the height indicates the number of radial sectors.

$$\hat{I}_{\log r, \theta} = \begin{bmatrix} \log r = \log \sqrt{(x^2 + y^2)} \\ \theta = \arctan \frac{y}{x} \end{bmatrix} \quad (1)$$

2.2 Eye representation in Log-Polar domain Figures 1(a)(b) show a set of eye images in Cartesian coordinate system and corresponding Log-Polar images. The eye images illustrate different scales, eye states translations and rotation angles. These differences are transformed into translations along θ and $\log r$ axes in the Log-Polar domain.

One can observe that a Log-Polar eye image can be divided into two areas; eyeball and skin. The eye contour is the boundary between the two areas and fully spans along θ axis. Consequently, detecting the eye contour is simplified to search on $\log r$ coordinates instead of searching on (x, y) coordinates in Cartesian coordinate system.

2.3 Eye color interpretation using PCA Detecting the $\log r$ coordinates of the eye contour must be achieved under different lighting conditions. The eyeball area which consists of iris, pupil and sclera can be represented approximately by fixed colors; black and white. These two colors are decomposed in RGB color space into (0,0,0) and (255,255,255); respectively. Therefore, the variance between the color channels in the eyeball area is very low whereas skin color cannot be represented by a fixed color and corresponding variance is considerably large.

In order to discern the variance between eye color channels, PCA is applied and the first eigenvector which indicates the direction of the most variation is used. Figure 1(c) shows the first eigenvector images of the images in Fig. 1(b). The pixel profiles (rows) in a first eigenvector image can uniformly be classified into two subsectors. Figure 1(d) represents an enlarged pixel profile that covered by a yellow line in Fig. 1(c). The first subsector may contain parts of sclera, iris and pupil. The corresponding pixel values are approximately small indicating low variance. The second subsector represents skin and it may contain eyebrow. The corresponding pixel values indicate high variance and change in an ascending grayscale order. The length of the two subsectors changes in a pixel profile according to the eye topological structure and θ with preserving the same order.

Figure 1: Eye image in Cartesian and Log-polar coordinate systems.

3. Eye contour detection

3.1 Detection strategy requirements Based on the uniform classification of pixel profiles in first eigenvector images, the eye contour can be extracted at the beginning of the second subsector which has an ascending order of pixel values. On the other hand, some noise may occur in the first subsector because of iris color or sclera red nerves. The noise appears with the same characteristics of the beginning of the second subsector as illustrated by the second left image in Fig. 1(c). In order to discriminate the second subsector and model the different occurrences of its beginning along $\log r$ axis, a pixel profile should fully be scanned with taking into account some previous scanned pixel values. Therefore, the detection strategy should have memorization property as well as generalization property. In order to meet these requirements, a recurrent neural network is used as detailed in the next sections.

3.2 Recurrent neural network Neural network can be classified into recurrent and non-recurrent networks [16]. In non-recurrent (feedforward) network, the output is calculated by propagating the input values through the feedforward connections. On the other hand, the recurrent networks use the previous input-output states as a part of current inputs. This makes recurrent networks more powerful to model time series signals [17] and more robust in terms of generalization and prediction [18].

The recurrent neural network consists of a set of neurons that are fully or partially connected to each other. The strength between neurons is expressed by a set of weights. Each neuron receives inputs from the other neurons and transmits its activation to the others as well.

Figure 2 shows the general structure of the recurrent network. It consists of three layers; input, hidden and output. The input layer receives the current input and oblique exter-

In order to train a recurrent network to model a function, training database and learning technique are used [19]. The error E between the desired output D and the actual network output O is minimized using Eq. (2). This can be optimized by updating the network weights using gradient descent algorithm and Chain rule [19]; Eq. (3).

$$\Delta w = -\eta \frac{dE(t)}{dw} \quad (3)$$

$$y_k(t) = \begin{cases} in_k(t) & \text{if } k \in I \\ O_k(t) & \text{if } k \in U \end{cases} \quad (4)$$
$$O_k(t+1) = f(net_k(t)) \quad (5)$$
$$\frac{\partial O_k(t)}{\partial net_k(t)} = O_k(t) * (1 - O_k(t)) \quad (7)$$
$$\frac{\partial \mathcal{O}_k(t+1)}{\partial w_{ii}} = \frac{\partial \mathcal{O}_k(t+1)}{\partial net_k(t+1)} \frac{\partial net_k(t+1)}{\partial w_{ii}} \quad (8)$$

$$\frac{\partial net_k(t+1)}{\partial w_{ij}} = \sum_{l \in U} w_{kl} \frac{\partial O_k(t)}{\partial w_{ij}} + \delta_{ik} y_j(t) \quad (9)$$

$$\Delta w_{ij} = -\eta \frac{\partial E(t)}{\partial O(t)} \frac{\partial O(t)}{\partial O_i(t)} \frac{\partial O_j(t)}{\partial w_{ij}} \quad (10)$$

$$\Delta w_{ij} = -\eta \frac{\partial E(t)}{\partial O(t)} \frac{\partial O(t)}{\partial w_{ij}} \quad (11)$$

$$O(t+1) = f(in(t), in(t-1), in(t-2), \dots, in(t-m), O_1(t), O_2(t), O_3(t), \dots, O_n(t)) \quad (12)$$

Each pixel profile contains a pixel/horizontal-segment of the eye contour as highlighted in Fig. 1(d). As long as the

Figure 3: Different pixel profiles.

pixel value changes smoothly in a pixel profile, the location of the eye contour is highlighted by creating a sudden change of the pixel value. This modified pixel profile is considered as a training profile and used to train the recurrent network to predict one pixel value ahead based on Eq. (12).

Accordingly, if a pixel profile is presented without any modification, the trained recurrent network is expected to predict the location of the eye contour by generating a sharp change of the pixel value at the actual location. Hence, an eye shape point is extracted. Repeating this operation for many pixel profiles in a first eigenvector image according to a fixed sampling rate of θ , a set of points are obtained. These points form the eye shape and considered as the detected eye contour.

4. Experimental results

4.1 Setup The proposed method, ASM and DT have been implemented in the environment of visual studio C++ 2008. Matlab has also been used to create some graphs.

Two databases have been used to evaluate the performance of the implemented methods. The first database includes 250 eye images from the UBIRIS database [20]; Fig. 4(a). The second database is locally collected to permit a considerable change of eye rotation with different scales and lighting conditions. The local database contains 100 right eye images as shown in Fig. 4(b). The eye images are resized to 150x100 for Cartesian coordinate system and 256x360 in Log-Polar domain. The eye contour is described by $H=16$ manually labeled points that form the eye shape.

In ASM, 150 manually labeled eye shapes have been used to model the dynamical change of the eye contour using Principal Component Analysis (PCA) [7].

In DT, the eye contour is described by two parabolas; the upper and lower eyelids. The two parabolas are sampled into 16 points to measure the accuracy. Four potential field images have been used to highlight the eye features, valley image highlights the iris, edge image highlights boundary of iris and eye contour, peak image highlights sclera areas and gray image which highlights the brightness inside iris (light reflectance); respectively.

In the proposed method, the recurrent network has the order of neurons of 15-40-1 for input, hidden and output layers; respectively. 50 Log-Polar eye images have been used to train the recurrent neural network by 3000 epochs. PCA is applied on RGB color channels of each training image and the corresponding first eigenvector image is obtained. The eigenvector image is rescaled between 0 and 0.9 pixel

(a) UBIRIS database[20].

(b) Local database.

Figure 4: Databases.

values and then transformed into a series of pixel profiles according to the fixed θ sampling rate $S=22$ rows. A pixel profile is 1X256 pixels and each image is represented by a series of $H=16$ pixel profiles. Consequently, a series consists of 16×256 pixel values.

4.2 Results In order to evaluate the performance of the proposed method; two training approaches have been considered. Figure 5(a) demonstrates the first approach by showing a pixel profile at $\theta=359$, green signal, taken from the left eye image in Fig. 1(c) and called raw signal. The signal is modified at the eye contour location to have a pulse with value of 0.8; blue signal. The modified signal is used to train the recurrent network in order to predict one future step. Once learning operation is done, the raw signal (without any modification) is presented to the trained network. The network predicts one value ahead and generates a pulse at the actual location of the eye contour; red signal.

Figure 5(b) shows the pixel profile at $\theta=355$ of the same eye image; green signal. This profile is presented to the

(a) Prediction of the same training eye profile $\theta=359$. (b) Prediction of different eye profile $\theta=355$.

Figure 5: Eye contour detection using the trained recurrent neural network.

Figure 6: Eye images represented by 16 eye profiles and eye contour detection using the trained recurrent network.

trained network and a predicted signal is generated. The predicted signal possesses a sharp pulse that expected to indicate the location of the eye contour; red signal.

Figure 6(a) illustrates the second approach of training the network by showing a series of $H=16$ pixel profiles; green signal. These profiles represent an eye image that sampled according to the sampling rate $S=22$ rows. The series is modified to have sharp pulses at the actual locations of the eye contour; blue signal. The modified series is considered as a training signal and used to train the network. After training, the raw series, which has no sharp pulses, is presented and a predicted signal is obtained; red signal.

Figure 7 shows some eye images taken from the databases. The eye contours are represented and manually labeled by $H=16$ points according to the sampling rate S as illustrated in red points. The predicted eye shapes by the trained recurrent neural network, based on the second training approach, are shown in green points.

Table 1 concludes the results of eye shape detection using the proposed method, ASM and DT. 300 eye images have been randomly taken from the two databases for this experiment. The table highlights three comparative components; detection time, accuracy and number of iterations.

5. Discussion

Besides the advantages of using Log-Polar transform to deal with different eye rotation angles, scales and translations as detailed in the previous work [10], PCA has provided a very stable eye texture representation against lighting conditions by highlighting the variance between RGB color channels. The representation provides a uniform description with reducing the nonlinearity of the change of pixel values along pixel profiles. As long as the pixel values change smoothly, a sharp pulse is guaranteed to be generated by the trained recurrent neural network one-time along a pixel profile.

The recurrent neural network performance has been investigated with respect to the change of pixel profiles (or θ coordinates). The recurrent network has been trained by a modified pixel profile as shown in Fig. 5(a). The trained net-

Figure 7: Eye shape detection using the proposed method. Red points are the exact eye shapes and green points are the detected eye shapes.

work was able to memorize the location of the eye contour when the raw pixel profile is presented. One can observe that there is a difference in magnitude between the training and predicted signals especially at the actual location of the eye contour. This difference doesn't affect the results because the most important information is the location of the rapid and sudden pulse in the predicted signal.

On the other hand, the difference between two pixel profiles in the same first eigenvector image is small with having the same features of classifications. Based on learning the raw profile in Fig. 5(a), the network was able to predict correctly the location of the eye contour in a pixel profile that is 6 degrees distant in Fig. 5(b).

However, the large difference in θ between the training and presented profiles influence the network performance by reducing the magnitude of the generated pulse. Therefore, the generated pulse becomes very difficult to be recognized in the predicted signals. Hence, training strategy has been changed to the second approach.

In order to make the network more flexible and robust to deal with different pixel profiles of various eye images, the network is trained using a series of modified pixel profiles. Each series represents an eye image according to the fixed sampling rate of θ as shown in Fig. 6. This strategy enables the network to discern various patterns of pixel profiles under different lighting conditions. Furthermore, it enables the network to learn the frequency of sharp pulses in each series that represent the eye shape in an image. Consequently, the changes of the eye contour under different eye states has been modeled by training the recurrent network using different series of pixel profiles. Therefore, the network performance has surprisingly been improved as shown in Fig. 6 and Fig. 7.

Figure 7 illustrates the performance of the proposed method in order to detect the eye contour of the testing images. The accuracy is high and the detected points almost match the manually labeled points.

The proposed method outperforms ASM and DT as high-

Table 1: Comparison between ASM, DT and the proposed method.

Method	Accuracy %	Iteration	Detection Time (msec)
ASM	69	20	85
DT	85	12	240
Prop. Method	94	-	59

lighted in Table 1. ASM and DT use an iterative strategy to detect the eye shape whereas the proposed method achieves fixed steps. This reduces the detection time significantly. The accuracy yielded is high because of using the prediction ability of the recurrent neural network as well as its generalization capabilities to deal with new pixel profiles.

One may ask about using feedforward network instead of recurrent network. In feedforward network, the patterns of data are not important whereas discerning the recursive relationship of the output with past states in the form of Eq.(12) is the objective of employing recurrent network. This dynamic form enables the proposed method to deal with noise effectively and to discriminate the eye contour significantly. On the other hand, the authors don't say that the current structure of the network is the best structure for modelling and predicting the eye contour location. We tried to use the same structure as a static network without any recurrent connections. The prediction ability was very low. Moreover, recurrent neural networks have widely been employed in neuro-control applications to be learned in real time. Therefore, the weights of the learned recurrent network can be updated online to track the eye contour in real time. This work is left to be investigated in future.

6. Conclusion

We proposed a method to detect the eye contour in grayscale eye images. The eye image is decomposed using Log-polar transform into a series of pixel profiles according to a fixed sampling rate. The pixel profiles are represented in a uniform classification by interpreting the variance between the eye color channels. The paper demonstrates a new approach for using recurrent neural networks in the object shape detection domain. A recurrent neural network is employed to achieve two tasks. The first task is in image domain to model the patterns of pixel values around the eye contour. The second achieved task is in shape domain to model the change of the eye contour at different eye states. Furthermore, the proposed method differs from other strategies by attaining fixed steps without using an iterative approach. The results verified that the proposed method demonstrates a very stable performance against the change of eye rotation angles, scales and eye states under different lighting conditions. High accuracy has been achieved with reducing the detection time compared to the conventional methods. Therefore, the proposed method is very promising to be used in real time applications.

References

- [1] D. O. Gorodnichy and G. Roth, "Use Your Nose as a Mouse", *International Conference on Vision Interface*, Vol.22, No.12, pp.354-361, 2002.
- [2] J. Zhu and J. Yang, "Subpixel Eye Gaze Tracking", *International Conference on Automatic Face and Gesture Recognition*, pp.131-136, 2002.
- [3] L. E. Sibert, R. J. K. Jacob and J. N. Templeman, "Evaluation and Analysis of Eye Gaze Interaction", *Journal of Modern Transportation*, NRL Report NRL/FR/5513-01-9990, Naval Research Laboratory, Washington, DC, 2001.
- [4] Q. Ji, Z. Zhu and P. Lan, "Real-time Nonintrusive Monitoring and Prediction of Driver Fatigue", *IEEE Trans. Vehicles Technology*, Vol.53, No.4, pp.1052-1068, 2004.
- [5] R. Barea, L. Boquete, M. Mazo and E. L6pez, "EOG guidance of a wheelchair using neural networks", *International Conference on Pattern Recognition*, Vol.4, pp.668-671, 2000.
- [6] T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active Shape Models-Their Training and Application", *Computer Vision and Image Understanding*, Vol.61, No.1, pp.38-59, 1995.
- [7] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active Appearance Models", *European Conference on Computer Vision*, pp.321-331, 1998.
- [8] A. L. Yuille, P. W. Hallinan and D. S. Cohen, "Feature Extraction from Faces Using Deformable Templates", *International Journal of Computer Vision*, Vol.8, No.2, pp.99-111, 1992.
- [9] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models", *International Journal of Computer Vision*, Vol.1, No.4, pp 321-331, 1988.
- [10] M. Aldibaja and S. Suzuki, "Eye Features Extraction Based on Eye Color Interpretation in Eigenspace and Eye Structure Representation in Logarithmic-Polar Domain", *Journal of Japan Society of Mechanical Engineering*, Vol.14, pp.264-271, 2014.
- [11] H. Araujo and J. M. Dias, "An Introduction to the Log-polar Mapping", *IEEE Second workshop on Cybernetic Vision*, pp.139-144, 1996.
- [12] A. Jackson, "A User's Guide to Principal Components", pp.1-25, 1991.
- [13] A. Giachetti, M. Campani and V. Torre, "The Use of Optical Flow for Road Navigation", *IEEE Transactions on Robotics and Automation*, Vol.14, No.1, pp.34-48, 1998.
- [14] A. Bernardino and J. Santos-Victor, "A Binocular Stereo Algorithm for Log-polar Foveated Systems", *Proceedings in Biologically Motivated Computer Vision*, Vol.2525, pp.127-136, 2002.
- [15] V. J. Traver and F. Pla, "The Log-polar Image Representation in Pattern Recognition Tasks", *Pattern Recognition and Image Analysis*, Vol.2652, pp.1032-1040, 2003.

- [16] M. N. Karim and S. L. Rivera, "Comparison of feed-forward and recurrent neural networks for bioprocess state estimation", *Computers and Chemical Engineering*, Vol.16, No.1, pp.369-377, 1992.
- [17] L. Medsker and L. C. Jain, "Recurrent Neural Networks: Design and Applications", *CRC Press*, 1999. ISBN 9780849371813
- [18] R. Williams and D. Zipser, "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks", *Neural Computation*, Vol.1, No.2, pp.270-280, 1989.
- [19] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Internal Representations by Error Propagation", *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol.1, pp. 312-366, 1986.
- [20] H. Proenca and L. Alexandre, "UBIRIS: A noisy Iris Image Database", *13th International Conference Image Analysis*, pp.970-977, 2005.

Mohammad Aldibaja (Member) was born in Lattakia, Syria, September, 1984. He received his PhD and MSc degrees from Toyohashi University of Technology. His research interests are in image processing applications and Human-Machine Interface Systems. Currently, he is a researcher in Toyohashi University of Technology.

Shinichi Suzuki (Non-member) is a professor of Toyohashi University of Technology, Japan. He received PhD degree from University of Tokyo with a major of Aeronautics. His research field is high-speed mechanics and optical measurement, in particular, non-equilibrium gas dynamics, rapid fracture phenomena and high-speed holography. Also interested in the generation of intelligence. He is member of JSEM, JSME and JSASS.