International Journal of Innovative
Computing, Information and Control ICIC International ©2011 ISSN 1349-4198
Volume 7, Number 7, July 2011 pp. **1--13**

Real Time Eyeball Tracking via Derivative Dynamic Time Warping for Human-Machine Interface

N. Mokhtar[1,]*, H. Arof[1] and M. Iwahashi[2]
[1]Department of Electrical Engineering
Faculty of Engineering, University of Malaya
50603 Lembah Pantai, Kuala Lumpur
Malaysia
*Corresponding author: norrimamokhtar@um.edu.my, norrimamokhtar.um@gmail.com

[2]Faculty of Engineering
Nagaoka University of Technology
Kamitomioka 1603-1, Nagaoka-Shi, Niigata
940-2188 Japan

Received March 2010; revised September 2010

ABSTRACT. *In this paper, a real time, user independent eyeball tracking approach is presented. The system is implemented using a low cost webcam. The robustness of the system is measured by several criteria such as users of different age where some of the users are wearing glasses under varying lighting condition, pose, eye orientation and distance from camera. The size and location of the region of interest which contains both eyes is made adaptive. Derivative Dynamic Time Warping is chosen as the classifier for this experiment since it can match patterns from data sequences with different lengths. Finally, the results, advantages, limitations and future works of the proposed method are reported. The online eye tracking procedure shows good accuracy and robustness when processing online image sequences at 50 frames/s on a 253 GHz Pavilion DV4 HP notebook.*

Keywords: Eyeball tracking, adaptive region of interest, Derivative Dynamic Time Warping, varying illumination condition.

1. Introduction. The disable community require special equipments to perform daily chores and other tasks. Human-computer interface (or interaction) technology may provide a means to assist them to lead a more independent life [1]. Common methods that realize human-machine interaction make use of brain, speech or visual signals as inputs to a computer or machine to perform specific tasks. For example, relative iris positions can be used as a cue to a vision system indicating which direction to go. In this paper, a real time eyeball tracking approach is presented wherein the position of the irises are located.

Table 1 summarizes some of the works involving facial feature recognition by comparing the databases, image processing techniques and classifiers used. In some of the works, images are obtained from well known facial image databases [2, 3, 4, 5, 6, 7, 8, 9], while in others, the images are captured online using compound eye imaging [10], cameras [11, 12, 13, 14] and CMOS digital imaging sensor [15]. Other approaches by Ohno et al. [16] used scanned images as input while Tsai [17] utilized captured gray scale images as compared to databases. For our proposed method, the input is acquired via online USB webcam.

In the majority of the methods in Table 1, the data (images) are treated as two dimensional (2D) arrays and thus 2D operations are performed on them [3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17]. However, in the work of Wijaya et al. [2], the image undergoes discrete cosine transformation and then the result is subjected to one dimensional analysis which consists of row and column operations. From Table 1, it can be observed that common image processing techniques are widely used together with statistical, neural and other types of classifiers. Apparently, each approach has its own advantages and requirements which are summarized in Table 2.

In this paper, a real time eyeball tracking system using DDTW (Derivate Dynamic Time Warping) is proposed. The system is user independent and is robust against minor variations in lighting and pose. The DDTW is chosen as our classifier due to its ability to perform identification despite variations in the sizes of the ROI (the region of interest which contains both eyes) and eyes. These variations occur when the distance between the user and the camera changes. Even when this distance is more or less fixed, the sizes of the ROI and eyes differ from person to person. The strength of the DDTW as a classifier stems from its ability to match two strings of one dimensional data of different length based on their pattern.

2. Methodology: DDTW (Derivative Dynamic Time Warping). DDTW is an improved version of the classic DTW (Dynamic Time Warping) method [18, 19]. While DTW matches two data sequences directly, DDTW obtains the first derivative of the data and then matches them. In this paper, only DDTW is described since it is used as the classifier of the system. Steps taken in the DDTW algorithm are shown in the flowchart in Figure 1.

Consider two data sequences X and Y of length n and m where x_i and y_j are the ith and jth elements of the first and second sequences respectively. To align the two sequences using DDTW we need to estimate the derivative of the two sequences first. At every data point (except endpoints), the derivatives are calculated as follows:

$$D_X(i) = ((x_i - x_{i-1}) + ((x_{i+1} - x_{i-1})/2))/2 \qquad (1)$$

$$D_Y(j) = ((y_j - y_{j-1}) + ((y_{j+1} - y_{j-1})/2))/2 \qquad (2)$$

TABLE 1. Image source, image processing techniques, preprocessing and classifier comparisons between existing works.

Authors [reference number] / Objectives	Image source / Database	Image processing techniques / pre-processing	Classifier
Proposed method / Eyeball detection	USB Webcam	Gray scale transformation, image subtraction, autohreshold, binarization, one dimensional conversion of the image information (rows and columns)	DDTW
Wijaya et al. [2] / Face recognition	ITS-Lab database, Kumamoto University, EE-UNRAM, Indian Face Database, Olivetti research laboratory(ORL)	Gray scale transformation, equalization, 1D-DCT analysis (rows and columns)	Extraction of dominant frequency features, multi-resolution metrics (similarity level based on statistical features, query frequency features and target frequency features)
Lin et al. [3] / Face recognition	CSU Face Identification Evaluation System	Geometric normalization, masking, histogram equalization and pixel normalization	PCA-Based SIFT with K-means algorithm
Zhou et al. [4] / Face recognition	Olivetti research laboratory(ORL) database	Discrete cosine transform, singular value decomposition	Hidden Markov Models – Support Vector Machine
Zhou et al. [5] / Image reconstruction for face recognition	Olivetti research laboratory(ORL) database, Yale face database, Georgia Tech face database	Centering, whitening, eigen-subspace	Fast ICA
Zheng et al. [6] / Eye features extraction	SJTU database (single eye image)	H channel of HSV colour space	Gabor eye-corner filter
Feng et al. [7] / Multi-cues eye detection	Face database from MIT AI laboratory	Snake algorithm, histogram to extract face region, erosion, dilation	Eye variance filter (potential eye window), Variance projection function (eye detection)
Li et al. [8] / Eye detection	AR face database, Rowley database	Histogram equalization, segmentation, edge detection, multithreshold	Similarity measure
Song et al. [9] / Eye detection	150 Bern images, 564 AR images	Multi-resolution wavelet transform, binary edge image, intensity	None
Miyazaki et al. [10] / Reconstruction of three dimensional image	Compound-eye imaging with defocus	Modified pixel rearrangement method for 3-D object, super-resolution algorithms	None
Kawato et al. [11] / Detection and tracking of eyes	Two cameras	Binarization, image differentiation	Geometry and pattern symmetry at face midpoint, template matching
Zhu et al. [12] / Eye detection	IR camera	Image subtraction, adaptive autothreshold, binarization	Support vector machine(SVM), Kalman pupil tracker
Bin et al. [13] / Face detection	USB camera	Log likelihood and threshold	Continuous Adaptative Mean SHIFT
Santis et al. [14] / Eye tracking	Low cost video camera	Binarization, four level segmentation, histogram for proper frame zoning	None
Amir et al. [15] / Eye detection sensor	CMOS digital imaging sensor	Image subtraction, threshold, find connected components, moments of components	None
Ohno et al. [16] / Content based image retrieval	24 scanned color images	Similarity measure	Correlation to reference images
Tsai [17] / Adaptive thresholding	300 captured grey scale images	Quadtree data structure, simulated annealing	Otsu thresholding

TABLE 2. Comparison of advantages and requirements of the proposed and the existing works.

Authors [reference number] / Objectives	Advantages	Requirements
Proposed method / Eyeball detection	Allow head movement, mouth movement, different background, user with and without glasses, different age, pose variation, eye orientation and distance variation	All users to focus at camera
Wijaya et al. [2] / Face recognition	Solved the retraining problem of PCA based face recognition, good performace	Same background for better performance comparison
Lin et al. [3] / Face recognition	Lower computational complexity, reduce dimension of feature space, robust to accessory and expression variation	Local descriptor, reservation of face region
Zhou et al. [4] / Face recognition	Reduce dimensionality and maintain facial features, alleviation of nonuniform illumination	None
Zhou et al. [5] / Image reconstruction for face recognition	Convenient for enlarging database, system's expansibility	Running time
Zheng et al. [6] / Eye features extraction	Color image	Estimated eye window is known
Feng et al. [7] / Multi-cues eye detection	Iris and eye corner locations detection	Limited hairstyles
Li et al. [8] / Eye detection	Eye and non-eye differentiation	Fuzzy template
Song et al. [9] / Eye detection	Variation in views and gaze directions	Users without glasses
Miyazaki et al. [10] / Reconstruction of three dimensional image	Defocus-blur restoration	Optical module, electronic signal processing module
Kawato et al. [11] / Detection and tracking of eyes	Allow head movement	Face midpoint test
Zhu et al. [12] / Eye detection	Successfully detect multiple face orientation	Same race with training image
Bin et al. [13] / Face detection	Moving face target	Background noise elimination
Santis et al. [14] / Eye tracking	Illumination change resistant	Image optimal segmentation, proper frame zoning
Amir et al. [15] / Eye detection sensor	Developed special hardware-based embedded system for eye detection	User without glasses
Ohno et al. [16] / Content based image retrieval	Simple	Subjective choice of the reference images
Tsai [17] / Adaptive thresholding	Lower misclassification error than Otsu thresholding	Additional computation

To align the two sequences using DDTW we construct an *n* by *m* matrix where the (*i*th, jth) element of the matrix contains the distance $d(D_X(i),D_Y(j))$ between the two points $D_X(i)$ and $D_Y(j)$ where *d* is the square of the difference between $D_X(i)$ and $D_Y(j)$. A warping path W, is a contiguous set of matrix elements that maps datapoints in X and Y. There are numerous possible mapping paths but we are only interested in the one that gives the minimum accumulated distance of the adjacent elements. Further constraints like continuity, monotonicity and other boundary conditions can be imposed on the warping paths depending on the applications.

The backtracking calculation will not be discussed since the information is not used in our eyeball tracking systems. However, the information is useful if reconstruction is essential. Further details on DDTW can be found in [18-19].

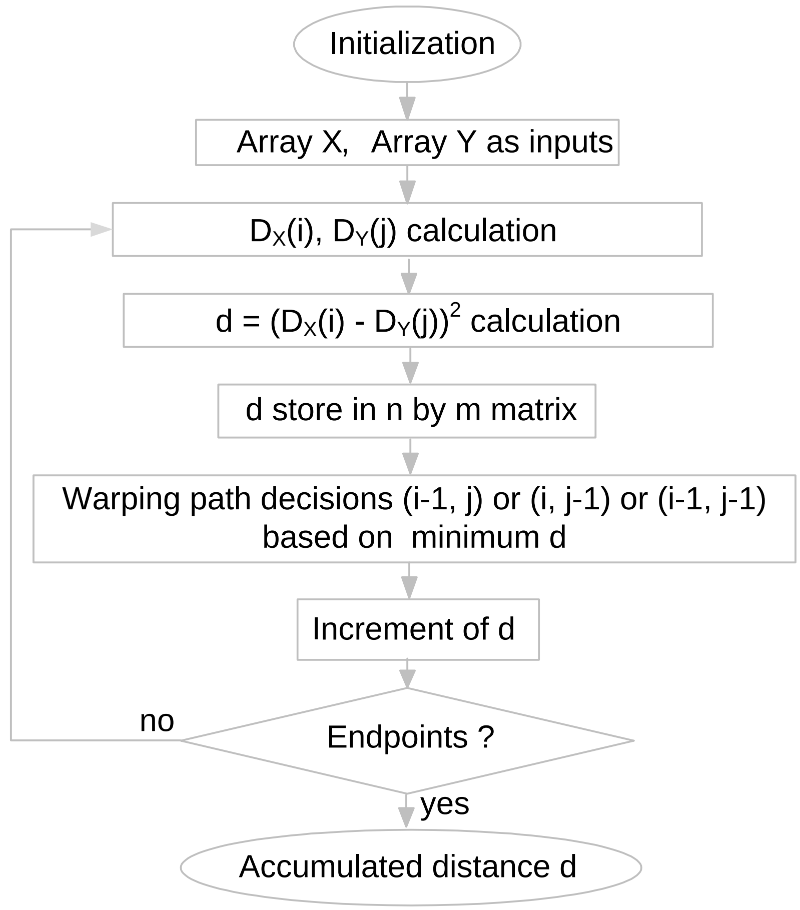


FIGURE 1. Flowchart of DDTW as a classifier.

2.1 Real time eyeball tracking algorithm. The flowchart for the real time eyeball tracking is shown in Figure 2. The entire program was developed using the LabVIEW Software (version 8.2) by National Instruments. The process starts with the USB webcam initialization. Once the image acquisition is done, the previous frame and the current frame of the images are kept in the memory buffer. These buffers are always updated. Then, image subtraction between the current frame and the previous frame is performed. Figure 3(a), shows the result of the subtraction process. From the subtracted image, binarization is done by segregating pixels with higher pixel values from those with lower values.

As can be seen in Figure 3(b), after binarization the eyes are marked as the white area. From this image, the intensities of pixels in the white areas are summed in row and column directions. The one dimensional arrays containing the sums of intensities of pixels in the white areas for each row and column are shown in Figure 4. They are called the online row and column arrays. These arrays are then matched with the template arrays (called the offline row and column arrays) using DDTW. If the offline and online arrays (which contain sequence of data) are matched, the calculated total distance d will be minimized as shown in Figure 4(a). Since our region of interest (ROI) is only the area that contains both eyes, we need to separate this area from other areas considered as the background. Thus we set a threshold such that if the online data represents the ROI, the obtained d (total minimum distance) should be lower than the threshold value. If d is less than the threshold value, then the **small motion detector** is triggered which indicates blinking eyes are detected as shown in Figure 4(b). The other motion such as movement in the background, head, mouth and nose movements will produce higher d (total minimum distance).

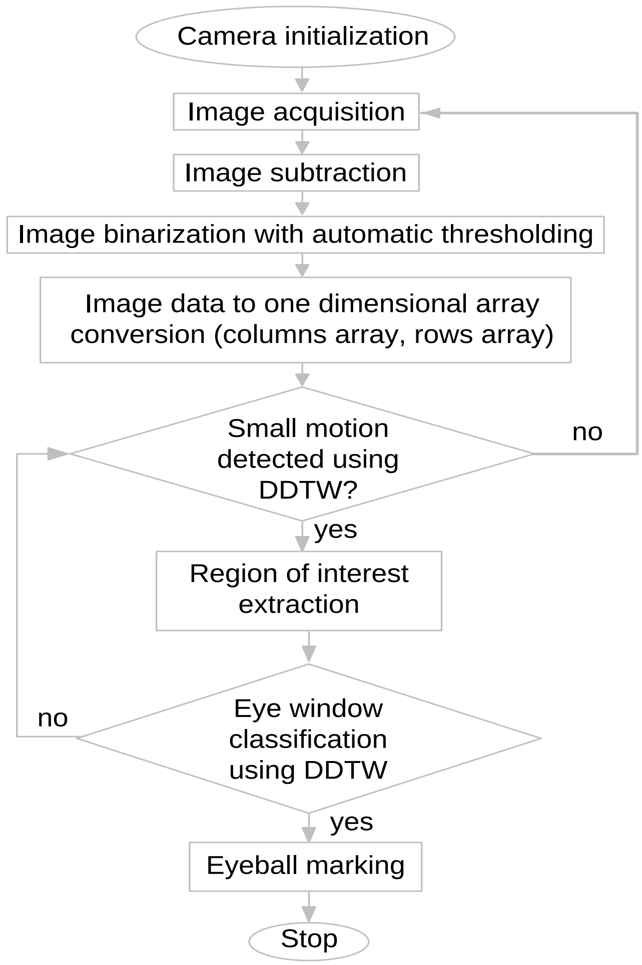
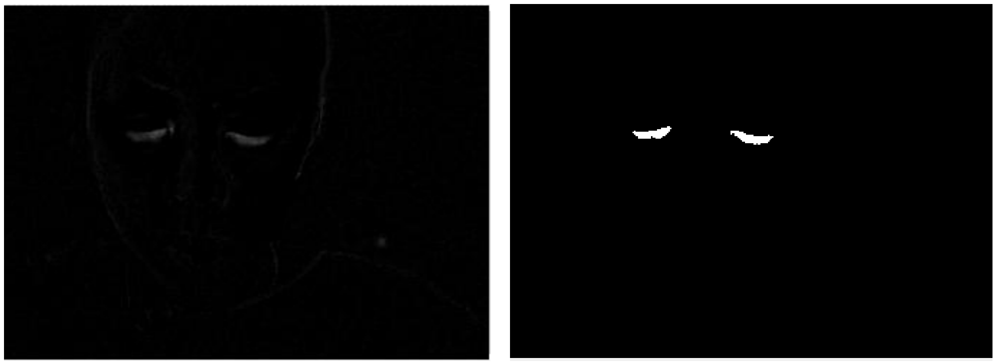


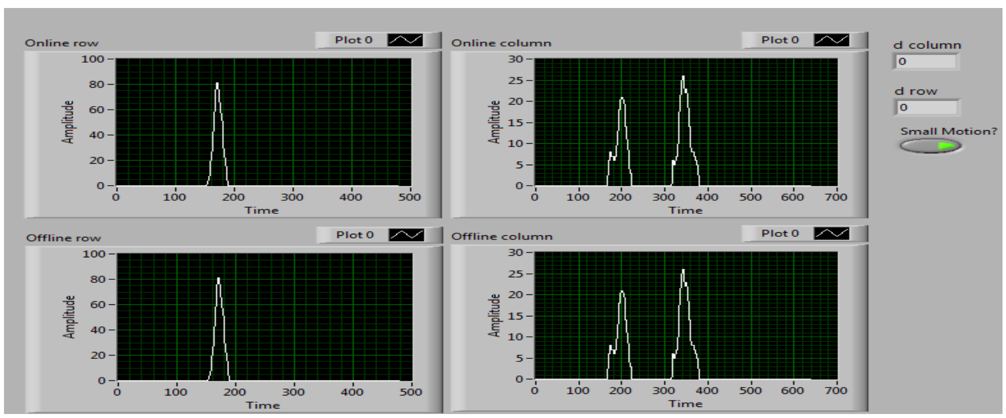
FIGURE 2. Flowchart of real time eyeball tracking.

After the small motion detector is triggered, the **region of interest** is set automatically based on the areas of the blinking eyes thresholded from the subtracted image. Then, from this region of interest, offline eye data and online eye data are matched using DDTW to classify the eyes. If both data are matched, **an adaptive eye window is displayed** on top of the input image. From the adaptive eye window, eyeball tracking is done via 'IMAQ find circles' function from LabVIEW Software (by National Instruments) morphology tools palette as shown in Figure 5. This process is continuous until the user presses the stop button.

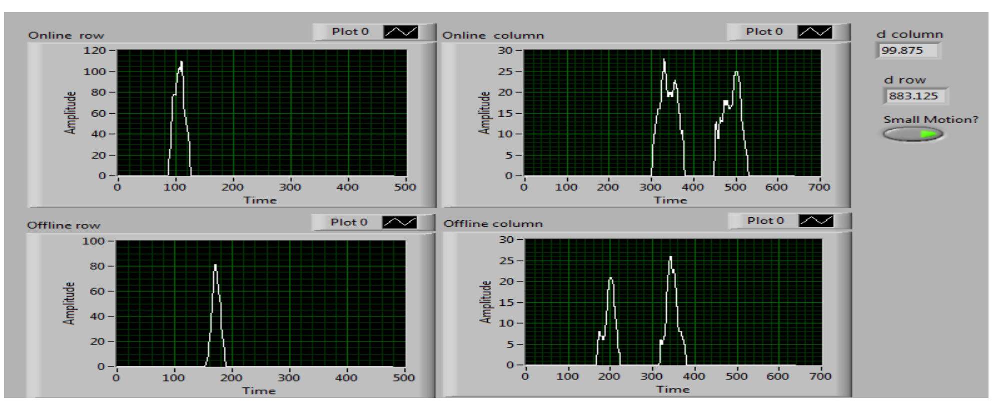


(a) (b)

FIGURE 3. (a) Image subtraction of previous and current frames, (b) Thresholded image of the substracted image.



(a)



(b)

FIGURE 4. Image data to row and column arrays conversion, d information; (a) Template and online data have equal data, d for column and d for row equal to 0; (b) Template and online data have different data, d for column = 99.875 and d for row = 883.125.

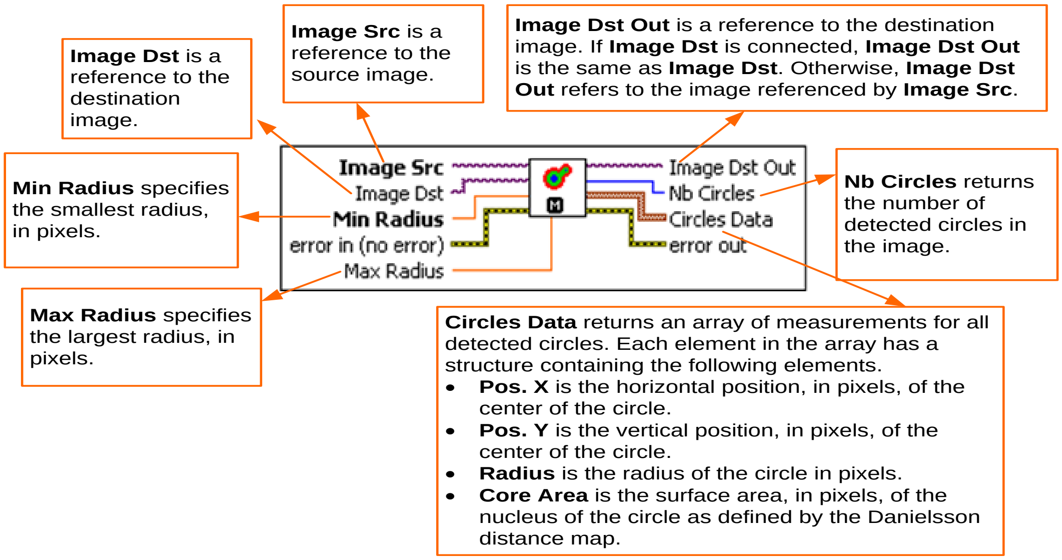


FIGURE 5. 'IMAQ find circles' function from LabVIEW Software (by National Instruments) morphology tools palette.

3. Experimental setup and conditions imposed to develop the main results. The experimental setup is shown in Figure 6. The webcam is wired by a USB cable to the notebook. It is assumed that the distance between the user and the webcam is within 50cm. Conditions that are necessary for the system to function properly are:

i. User should be seated in front of the USB webcam.
ii. User should reduce movement while triggering the small motion detector to initiate the system.
iii. User should focus and look into the camera until the eye window is detected and displayed.
iv. To re-initialize the system, the user can make obvious movement which caused the small motion detector to flag 'false condition'.

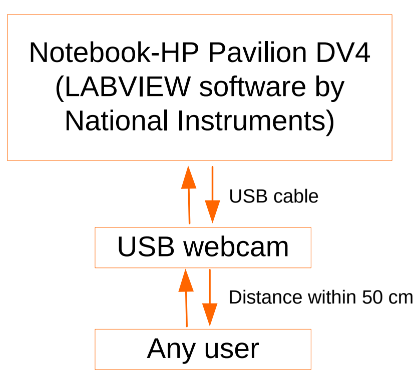


FIGURE 6. Experiment conditions.

4. Results and discussion. The effectiveness of real time eyeball tracking system is measured by:

i. Users (with glasses and without glasses)
ii. Lighting condition (illumination change)
iii. Pose (turn right, turn left and face down)
iv. Eye orientation
v. Distance
vi. Different age

Figure 7 shows that from the adaptive region of interest window setting, only the eye area is extracted. This eye segment is then converted into gray scale, binarization and finally into one dimension arrays so that it can be processed via DDTW with less computational complexity.

Figure 8, depicts an example of the computation time taken for eye window detection at 0.083 second. After the eye window is detected, the computation time for eyeball detection is around 0.216 second. The time taken for eyeball detection is slightly longer owing to the graphic added to the image such as the caption 'Eyeball detected!' and red circle as the marker to the eyeball. The result shows that the total computation time for eye window tracking and eyeball detection is 0.299 second which is reasonable for real time response.

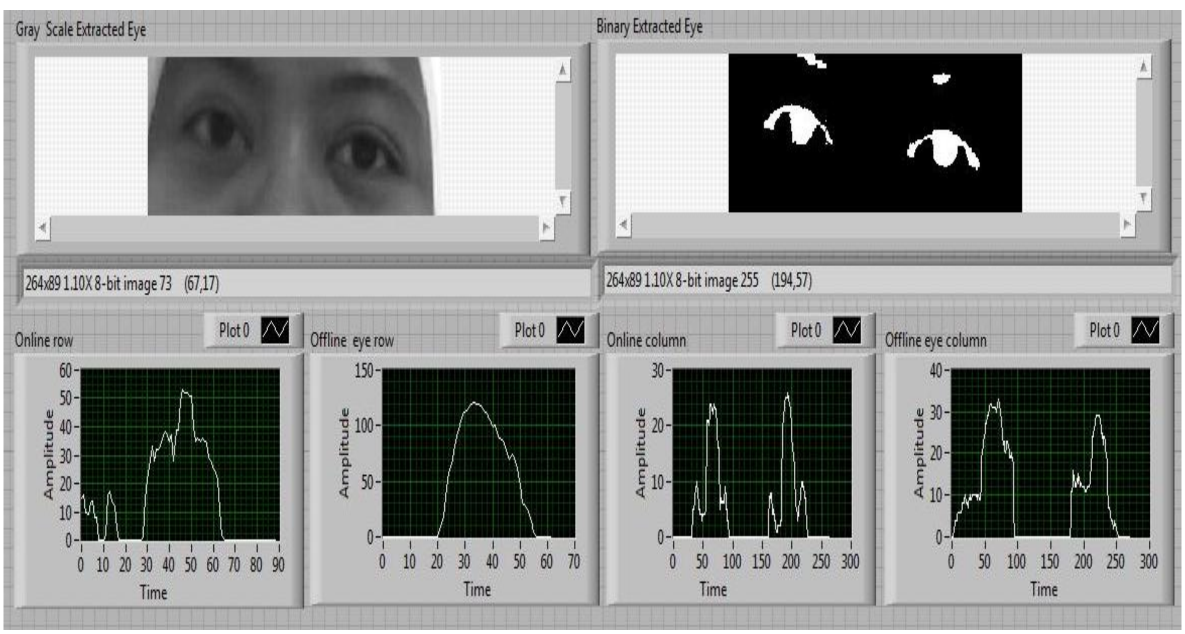


FIGURE 7. Adaptive region of interest with the extracted image in gray scale, binarization of image, offline arrays and online arrays.

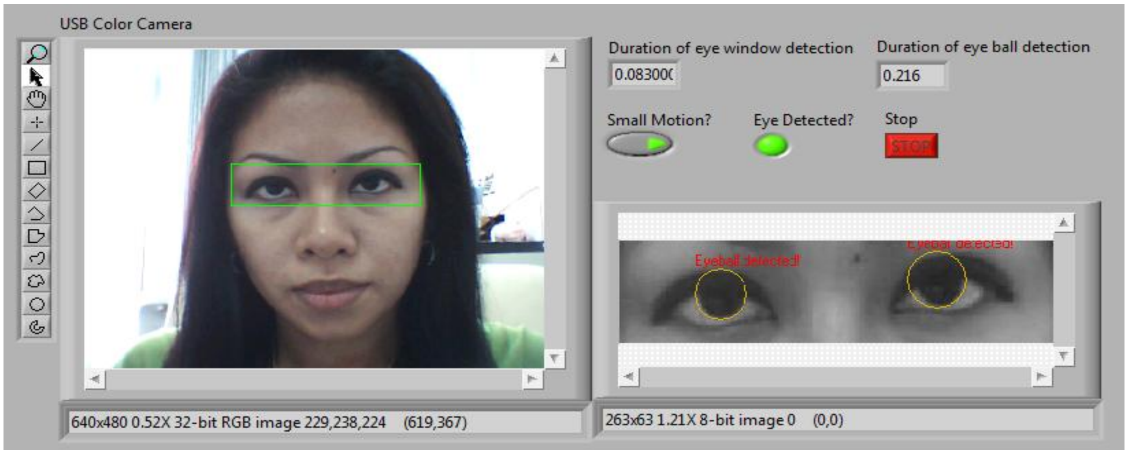


FIGURE 8. Computation time for eye window tracking and eyeball detection (in second).

Figure 9 displays the effectiveness of the approach which are measured based on users both with glasses and without glasses, varying lighting condition, pose variations, eye orientation, different age and varying distance. The eyeball tracking system worked well with multiple users, for both whether the user is wearing glasses and otherwise. The system also worked well with multiple ages for both adults and children. The robustness of the system was tested by conducting the experiments under pose variations such as left pose, right pose and faces downward. The spacing criterion was also tested so that the user did not have to be very near to the camera hence the user becomes comfortable while using it. The results show that DDTW provides good results under various conditions tested. The illumination problems have been overcome by auto-threshold method which was described in the previous section.

Despite its effectiveness, the system has some limitations such as inability to detect user with very small eyes and very dark eye bag. The system is incapable to withstand vibration if the system is mounted on a moving vehicle.

5. Conclusions. A practical eyeball tracking system should be robust, comfortable to use, cheap and fast enough for real time application. The proposed method has been shown to possess all these attributes. The effectiveness of the approach was measured against users from different ages. Some of them are wearing glasses. The system was also tested under varying lighting conditions, pose variations, eye orientations and camera spacing. The method capitalizes on the flexibility of DDTW to classify the image data based on the shape of the arrays rather than the size of the arrays. By implementing the adaptive region of interest for image extraction, the size of data for classification is reduced to enable faster computation for real time response of the system.

For future works, the system should be refined to increase the accuracy of recognition on users with small eyes and very dark eye bag. When the system is mounted on a moving vehicle, vibration causes the captured image to become blurred. This in turn decreases the recognition rate. These issues need to be addressed before the system can be fully integrated as an interface between human and machine.

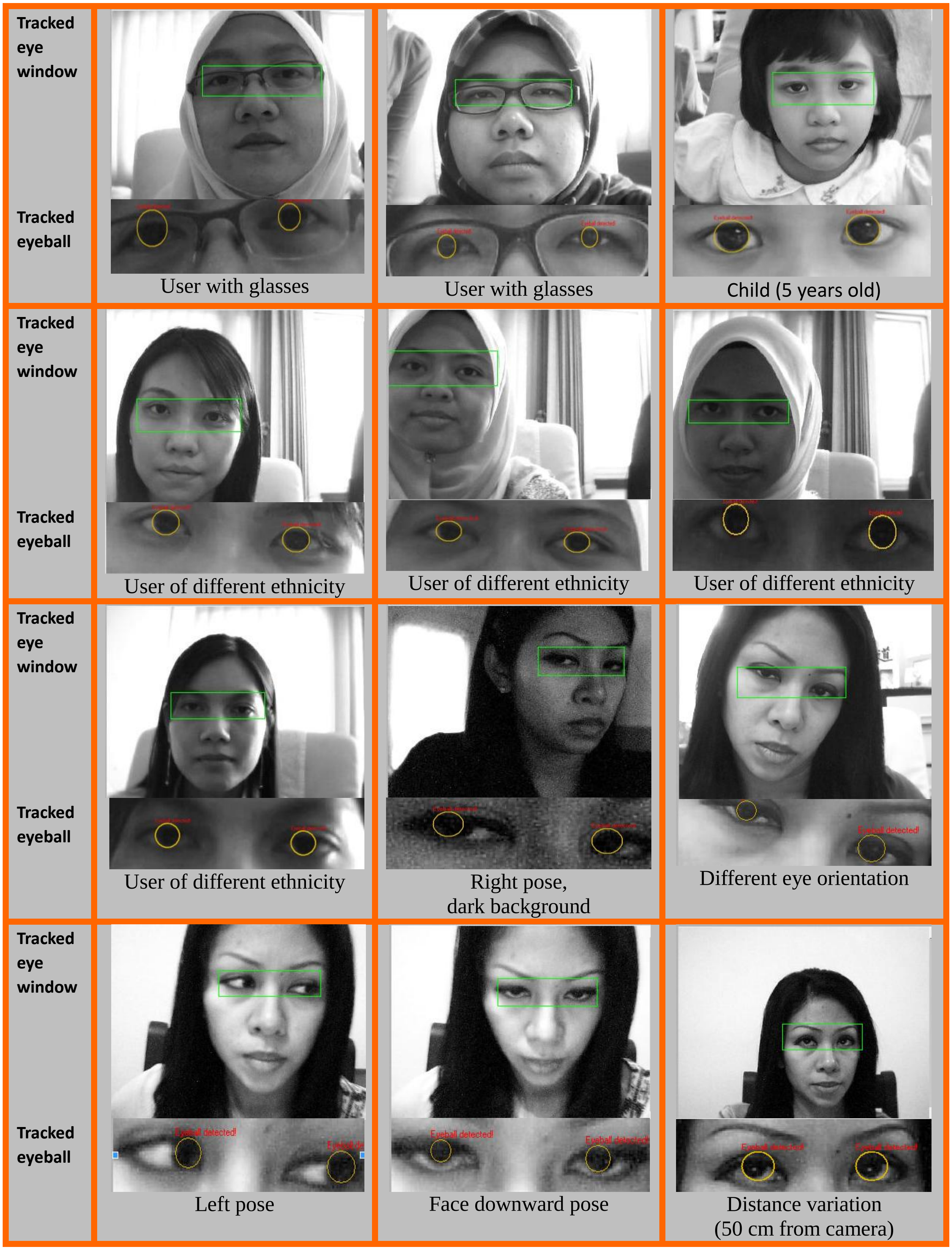


FIGURE 9. Results of the real time **eyeball tracking** in various conditions.

Acknowledgment. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] F. Wang, X. Ren, A survey on human computer interaction technology for disable persons, *International Journal of Innovative Computing, Information and Control,* vol.6, no.6, pp.597-608, 2010.

[2] I. G. P. S. Wijaya, K. Uchimura, Z. Hu, Face recognition based on dominant frequency features and multiresolution metric, *International Journal of Innovative Computing, Information and Control,* vol.5, no.3, pp.641-651, 2010.

[3] S. D. Lin, J. H. Lin, C. C. Chiang, Combining scale invariant feature transform with principal component analysis in face recognition, *ICIC Express Letters,*vol.3, no.3, pp.1-6, 2009.

[4] C. Zhou, X. Wei, Q. Zhang, Face recognition based on HMM-SVM, *ICIC Express Letters,* vol.1, No.1, pp.1-6, 2007.

[5] C. Zhou, X. Wei, Q. Zhang, B. Xiao, Image Reconstruction for face recognition based on fast ICA, *International Journal of Innovative Computing, Information and Control,* vol.4, no.7, pp.1723-1732, 2008.

[6] Z. Zheng, J. Yang, L. Yang, A robust method for eye features extraction on color image, *Pattern Recognition Letter*, vol.26, pp.2252-2261, 2005.

[7] G. C. Feng, P. C. Yuen, Multi-cues eye detection on gray intensity image, *Pattern Recognition Letter,* vol.34, pp.1033-1046, 2001.

[8] Y. Li, X. L. Qi, Y. J. Wang, Eye detection by using fuzzy template matching and feature-parameter-based judgment, *Pattern Recognition Letters*, vol. 22, pp.1111-1124, 2001.

[9] J. Song, Z. Chi, J. Liu, A robust eye detection method using combined binary edge and intensity information, *Pattern Recognition*, vol. 39, pp.1110-1125, 2006.

[10] D. Miyazaki, K. Ito, Y. Nakao, T. Toyoda, Y. Masaki, Reconstruction of three dimensional image from compound-eye imaging with defocus using ray tracing, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11B, pp.4225-4235, 2009.

[11] S. Kawato, N. Tetsutani, Detection and tracking of eyes for gaze-camera control, *Image and Vision Computing*, vol. 22, pp.1031-1038, 2004.

[12] Z. Zhu, Q. Ji, Robust real-time eye detection and tracking under variable lighting conditions and various face orientations, *Computer Vision and Image Understanding*, vol.98, pp.124-154, 2005.

[13] A. S. K. Bin, L. Y. Kang, Face detection and tracking utilizing enhanced CAMSHIFT model, *International Journal of Innovative Computing, Information and Control,* vol.3, no.3, pp.597-608, 2007.

[14] A. D. Santis, D. Iacoviello, Robust real time eye tracking for computer interface for disable people, *Computer Methods and Programs in Biomedicine*, vol. 96, pp.1-11, 2009.

[15] A. Amir, L. Zimet, A. S. Vincentelli, S. Kao, An embedded system for an eye-detection sensor, *Computer Vision and Image Understanding*, vol. 98, pp.104-123, 2005.

[16] A. Ohno, H. Murao, A similarity measuring method for images based on the feature extraction algorithm using reference vectors, *International Journal of Innovative Computing, Information and Control*, vol.5, no.3, pp.1-9, 2009.

[17] Y. H. Tsai, A new window selection for local image thresholding under uneven illuminations, *International Journal of Innovative Computing, Information and Control,* vol.6, no.3A, pp.1059-1067, 2010.

[18] X. Huang, A. Acero, H. W. Hon, *Spoken language processing, A guide to theory, algorithm and system development,* Prentice Hall, 2001.

[19] E. Keogh, M. Pazzani, Derivative dynamic time warping, *First SIAM International Conference on Data Mining*. Chicago, Illionis,pp.1-11, 2001.

[20] L. R. Rabiner, C. Schmidt, Application of dynamic time warping to connected digit recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. assp-28, no.4, pp.377-388,1980.

[21] M. K. Brown, L.R. Rabiner, An adaptive, ordered, graph search technique for dynamic time warping for isolated word recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. assp-30, no.4, pp.535-544, 1982.