Scientific Research and Essays Vol. 5 (19), pp. 2947-2952, 4 October, 2010
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 ©2010 Academic Journals

Full Length Research paper

One dimensional image processing for eye tracking using derivative dynamic time warping

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

Accepted 14 July, 2010

In this paper, one dimensional image processing for eye tracking approach is presented using a low cost webcam. It was found that the algorithm is resistant to illumination variation and is able to track eyes of users with and without glasses in multiple orientations. The size and location of the region of interest (ROI), which contains both eyes, is adaptive. Derivative Dynamic Time Warping (DDTW) is chosen as the classifier for this experiment since it can match patterns from one dimension data sequences with varying length. The DDTW is applied to compare distances of the area with an offline template by investigating separately, the row arrays and column arrays for eye tracking. The online one dimensional image processing for eye tracking procedure shows good accuracy and robustness when implemented at 50 frames/s on a 253 GHz Pavilion DV4 HP notebook.

Key words: Eye tracking, adaptive region of interest, derivative dynamic time warping, illumination variation resistant, one dimensional image processing.

INTRODUCTION

One of the applications of human machine interface tools is to facilitate the needs of the disable community. The main objective of the research in this area is to produce suitable facilities that would enable them to live an independent and productive life. The better the facilities, the easier it is for them to become more independent in managing their own life. An example of the human machine interface is the brain computer interface which allows the disabled to control the wheelchair via EEG (Electroencephalography) signals (RIKEN, 2010). However, such a system usually requires specialized sensors and electronic interface. It is widely noted that its cost and cumbersome initial preparation, prohibit its widespread application.

A cheaper alternative is to track the movement of the eyes as an input signal which determines the direction of the motorized wheelchair. Besides its low cost, the system offers robustness, comfort and real time response. Considering all these factors, many researches focus on eye tracking as one of the methods to achieve human machine interface. Eye tracking systems are cheaper system compared to brain computer interface systems. By improving the methods and algorithm, the robustness of eye tracking systems can be improved. It also provides comfort to the user since fewer wires are involved and there is no gels involvement as compared to brain computer interface systems. Like the current brain computer interface systems, eye tracking systems can also provide real time response.

From these motivations, the eye tracking systems have been conducted. Some researchers conducted offline eye detection which used the images from database (Zheng et al., 2005; Feng and Yuen, 2001; Wang and Ji, 2007). For the human machine interface application, the eye tracking need to be conducted in real time situation in order to control the movement of any devices. Kawato's method is also good and reliable (Kawato and Tetsutani, 2004). They used two cameras and head movement cancellation for real time eye detection. Head movement during eye detection was allowed. Some of the researchers also developed eye tracking method using active IR illumination sensor and special embedded systems together with object recognition techniques (Zhu and Ji, 2005; Amir et al., 2005).

*Corresponding author. E-mail: norrimamokhtar@um.edu.my.

From the image processing point of view, Santis's approach relied on frame zoning and optimal segmentation algorithm, although no eye movement model is required (Santis and Iacoviello, 2009). The frame zoning method requires the user's face to be in the centre of the frame for proper zoning. For further development, Santis also suggested the need to consider subjects wearing glasses and changing background. Some of the eye detections also utilized binarization, edge detection, locating skin region, face region extraction and background subtraction techniques (Zheng et al., 2005; Feng and Yuen, 2001; Kawato and Tetsutani, 2004; Zhu and Ji, 2005; Amir et al., 2005; Santis and Iacoviello, 2009; Li et al., 2001; Song et al., 2006).

From the eye detection classifier point of view, variance projection function was used with offline eye detection and plain background (Feng and Yuen, 2001). Wang's approach, which used the Recursive Nonparametric Discriminant Analysis (RNDA) combined with AdaBoost techniques for multi-view face detector, was also conducted from image database or offline (Wang and Ji, 2007). Kawato's method used the template matching technique that was triggered by blink detection (Kawato and Tetsutani, 2004). This method also used three geometrical conditions to classify eye detection. Zhu's method used two combinations of classifier, which are Kalman filter and support vector machine to complement each of these methods' weaknesses (Zhu and Ji, 2005). The input for this method was obtained from infrared images. Li's method involves the use of Fuzzy template matching (Li et al., 2001). The judgement of eye and non-eye is based on similarity measures between the input image and the template. The similarity measures are defined by five parameters where the thresholds are obtained from experimental data. The illumination problem was overcome by multiple threshold setting of the image.

In this paper, real time independent user eye tracking systems using DDTW (Derivate Dynamic Time Warping) have been conducted. Based on the developed methods in Zheng et al. (2005), Feng and Yuen, (2001), Wang and Ji, (2007), Kawato and Tetsutani, (2004), Zhu and Ji, (2005); Amir et al. (2005), Santis and Iacoviello, (2009), Li et al. (2001) and Song et al. (2006), auto threshold, binarization, automatic or unfix region of interest, template model and classifier need to be done in order to achieve good results of real time eye tracking systems. DDTW has been chosen as our classifier. The reasons for the decision of using DDTW are due to its ability to:

1. Adapt with variable size of eyes
2. Adapt with variable distance
3. Unfix and automatic region of interest
4. Classify based on shape of the data

Previously, DDTW was used in multiple signal processing research (Huang et al., 2001; Keogh and Pazzani, 2001; Rabiner and Schmidt, 1980; Brown and Rabiner, 1982; Karungaru et al., 2007; Bin et al., 2007). The current trend in image processing technique is to process the image data in two dimensions of arrays of pixel values. In order to simplify the DDTW application in real time image processing, the image information is transformed into two separate arrays containing row and column information of the pixel values. From this transformation, the warping process via DDTW was done in this research.

METHODOLOGY

DDTW (Derivative Dynamic Time Warping)

DDTW was an improvement of DTW (Dynamic Time Warping) method (Huang et al., 2001; Keogh and Pazzani, 2001). However, in this paper, only DDTW is described because it is used as the classifier of the systems. DDTW provides information about the shape of the arrays by considering the first derivative of sequences. DDTW attempts to align two sequences based on the shape of the feature. The algorithm details of the applied DDTW are displayed in the flowchart shown in Figure 1a. In this work, the flow of the DDTW begins with the initialization of the function within the main program. The initialization begins with total minimum distance D which is given as:

$$D(1,1) = d(1,1) = 0 \tag{1}$$

where d is the distance between two points q_i and c_j. The q_i and c_j are points that belongs to array $x(i)$ and $y(j)$. These arrays have i and j as their elements. In order to calculate D_x of array $x(i)$ and D_y of $y(j)$:

$$D_x(q) = ((q_i - q_{i-1}) + ((q_{i+1} - q_{i-1})/2))/2) \tag{2}$$

$$D_y(c) = ((c_j - c_{j-1}) + ((c_{j+1} - c_{j-1})/2))/2) \tag{3}$$

The D which is the total minimum distance is given by the summation of:

$$D = (D_x(q_i) - D_y(c_j))^2 \tag{4}$$

The backtracking calculation will not be discussed because the information is not used in eye tracking systems. However, the information is useful if reconstruction is required in certain project. After the calculation of the total minimum distance D, from the flowchart shown in Figure 1a, the warping path at a certain point q_i and c_j is set to either $(i\text{-}1,j)$, $(i, j\text{-}1)$ or $(i\text{-}1,j\text{-}1)$ direction. The warping decision direction is determined by the direction that provides the minimum D. The total minimum distance of D is updated from time to time. The program is looped by the size of the arrays.

Real time independent user eye tracking algorithm

In order to achieve real time with independent user application, a simple and robust classifier needs to be chosen correctly. In this case, DDTW has been chosen due to its capability to classify, based on the shape of the one dimension arrays and can match patterns from data sequences with varying length (Keogh and Pazzani, 2001).

Based on the developed function of the classifier, the procedure is summarized as shown in Figure 1b. The whole program was developed by using LabVIEW Software (version 8.2) by National Instruments.

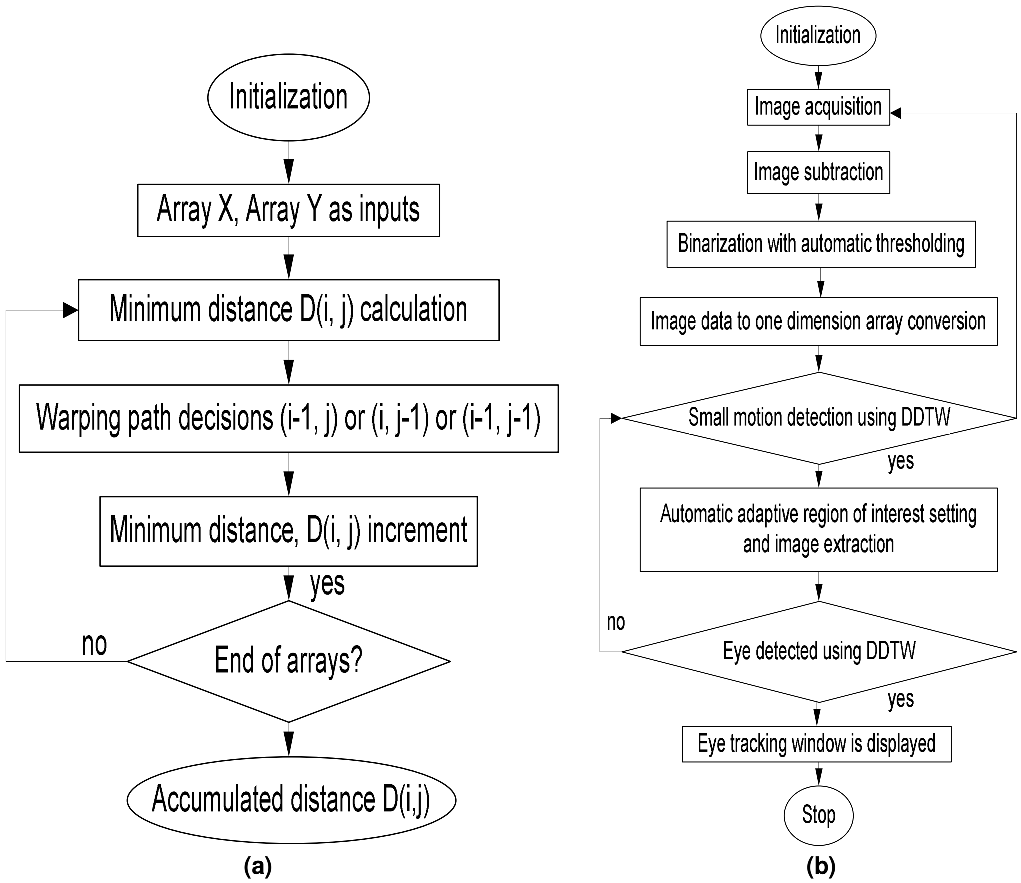

Figure 1. (a) Flowchart of DDTW as classifier, (b) Flowchart of real time independent user eye tracking algorithm.

The process starts with USB webcam initialization. After the image acquisition was done, the previous and current frames of the images were kept in the memory buffer. These buffers were always updated. Then, image subtraction between the current frame and previous frame was done. From the subtracted image, the autothreshold process was done via selecting the maximum pixel value since the difference between the two images will produce higher pixel values.

After the subtracted image has been thresholded, the moving object in the image is marked as white. From this image, the pixel values are converted into row arrays and column arrays, which is shown in Figure 2. From these arrays, by referring to DDTW, the D (total minimum distance) between row, column arrays of online and offline data are calculated. The offline data which consist of row array and column array are used as template. If the offline data and online data are equal, the calculated D (total minimum distance) is equal to zero. If the offline data and online data are blinking eye, the obtained D (total minimum distance) should be within the threshold value. If D (total minimum distance) is less than the threshold value, then the small motion detector is triggered. In this case, the small motion detector is only triggered when eye blinking is detected. The other motion such as background or clear head movement will give a higher D (total minimum distance).

After the small motion detector was triggered, the automatic region of interest was set based on the window of the blinking eye from the threshold subtracted image. Then, based on the variable or unfix region of interest, offline eye data and online eye data was classified using DDTW. Similarly, if the online data is not an eye which is either nose, mouth or jaw movement, *D* (total minimum distance) will also give a higher value than the threshold value. If the 'eye is detected', where the *D* (total minimum distance) is within the threshold value, an 'eye window is displayed' on top of the input image.

Small head movement is well compensated as long as the small motion detector is true. This process is continuous until the user presses the stop button.

Experimental setup

In the experiment, a Logitech webcam was wired to the notebook via USB cable. The distance between the user and the webcam must be less than 50 cm.

RESULTS AND DISCUSSION

The user eye tracking system is tested with variations in:

1. Users (with glasses and without glasses)
2. Lighting condition (illumination change)
3. Pose (turn right, turn left and downward)

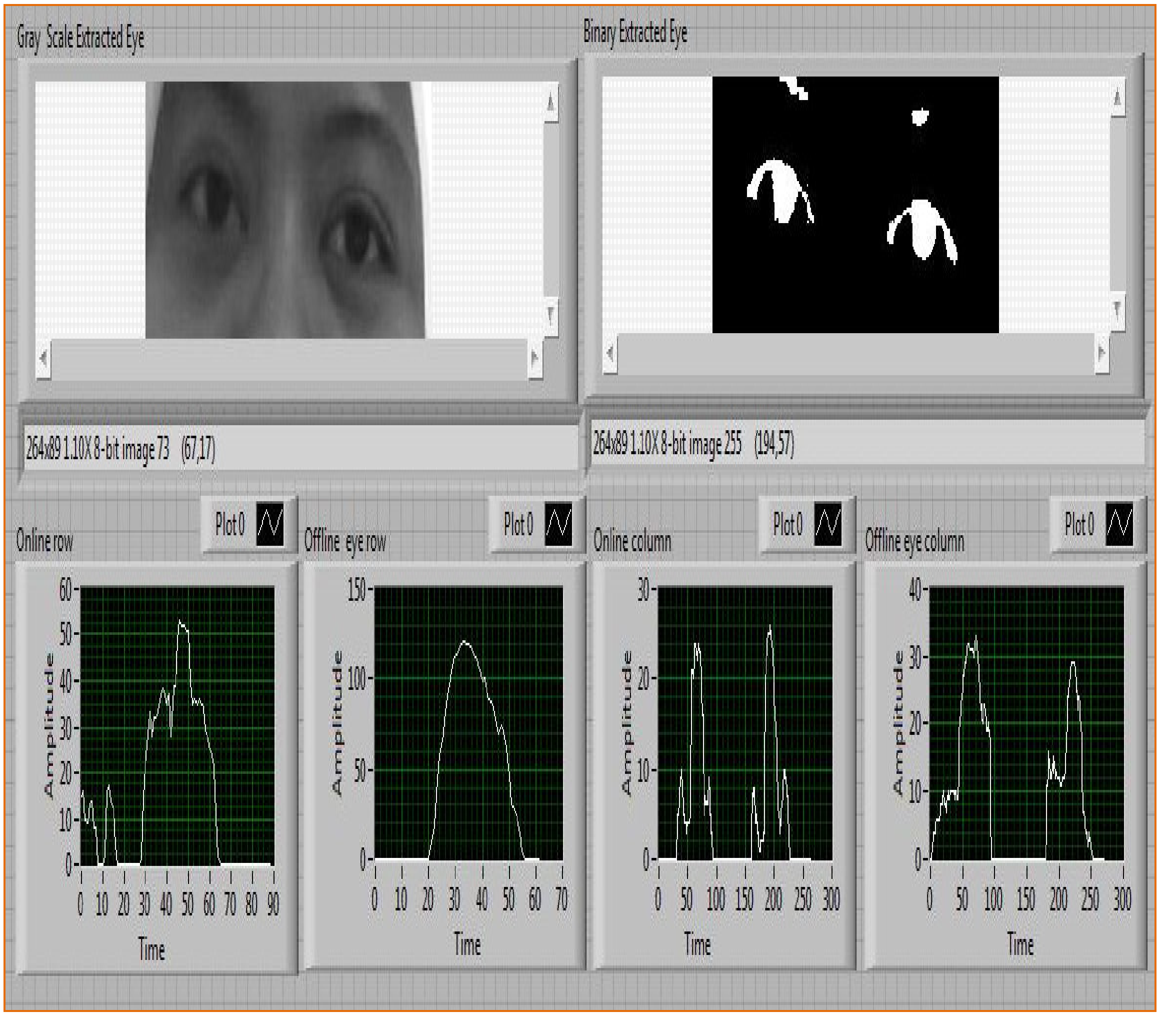

Figure 2. Unfix or variable region of interest with the extracted image in grayscale, binarization, offline column arrays and online row arrays.

4. Distance

Figure 2 shows that from the selected 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.

From the results shown in Figure 3, we can deduce that the one dimensional image processing for eye tracking system works well with multiple users, with glasses and without glasses. The system also works on children. The robustness of the system was tested when experiments was conducted under multiple poses which included left pose, right pose and downward pose. The distance variation was also tested so that the user does not have to be very close 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 autothreshold method, which was described previously.

Conclusions

A good eye tracking systems should be robust, provide comfort to user, low cost and real time responses. Based on the results, the developed real time independent user eye tracking systems provide good results in accordance with these criteria. The methods made use of flexibility of DDTW to classify the image data based on the shape of the arrays rather than the size of the arrays. The robustness of the system was also tested under various conditions such as multiple users (with glasses and without glasses), multiple ages, multiple lighting conditions (illumination change), multiple poses (right pose, left pose and downward pose) and variation of distances. By implementing the adaptive region of interest for image

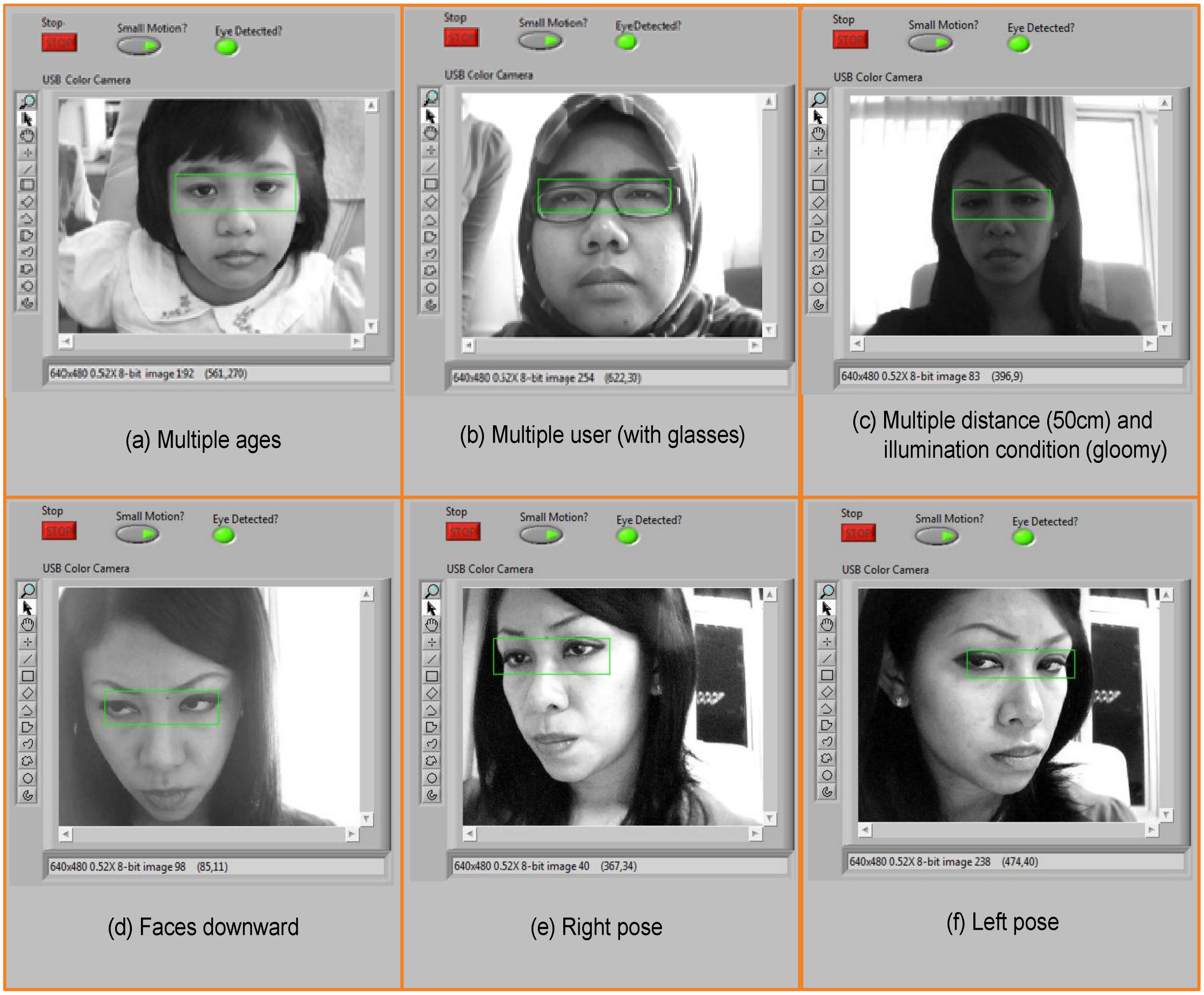

Figure 3. Results of the real time independent user eye tracking in various conditions.

extraction, the size of arrays (which are converted to one dimensional arrays) for classification has been reduced, hence it provides faster computation time and enable real time application. In the next project, eye ball tracking will be considered to provide effortless human-machine interface application.

REFERENCES

Amir A, Zimet L, Vincentelli AS, Kao S (2005). An embedded system for an eye-detection sensor. Computer Vision and Image Understanding, 98: 104-123.

Santis AD, Iacoviello D (2009). Robust real time eye tracking for computer interface for disable people. Comput. Methods Programs Biomed., 96: 1-11.

Bin ASK, Kang LY (2007). Face detection and tracking utilizing enhanced CAMSHIFT model. Int. J. Innovative Comput. Inf. Control, 3(3): 597-608.

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

Feng GC, Yuen PC (2001). Multi-cues eye detection on gray intensity image. Patt. Recog. Lett., 34: 1033-1046.

Song J, Chi Z, Liu J (2006). A robust eye detection method using combined binary edge and intensity information. Patt. Recog., 39: 1110-1125.

Rabiner LR, Schmidt C (1980). Application of dynamic time warping to connected digit recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, assp-28(4): 377-388.

Brown MK, Rabiner LR (1982). An adaptive, ordered, graph search technique for dynamic time warping for isolated word recognition.

IEEE Transactions on Acoustics, Speech and Signal Processing. assp-30(4): 535-544.

Wang P, Ji Q (2007). Multi-view face and eye detection using discriminant features. Computer Vision and Image Understanding, 105: 99-111.

RIKEN (2010). Toyota Motor Corporation (2010). Real – time control of wheel chair with brain waves. Science Daily. http://www.sciencedaily.com/ releases/2009/06/090629101848.htm.

Karungaru S, Fukumi M, Akamatsu N (2007). Automatic human faces morphing using genetic algorithms based control points selection. Int. J. Innovative Computing, Inf. Control, 3(2): 247-256.
Kawato S, Tetsutani N (2004). Detection and tracking of eyes for gaze-camera control. Image and Vision Computing, 22: 1031-1038.
Huang X, Acero A, Hon HW (2001). Spoken language processing, A guide to theory, algorithm and system development. Prentice Hall.
Li Y, Qi XL, Wang YJ (2001). Eye detection by using fuzzy template matching and feature-parameter-based judgement. Patt. Recog. Lett., 22: 1111-1124.
Zheng Z, Yang J, Yang L (2005). A robust method for eye features extraction on color image. Patt. Recog. Lett., 26: 2252-2261.
Zhu Z, Ji Q (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. Computer Vision and Image Understanding, 98: 124-154.