ELSEVIER

Optics & Laser Technology 34 (2002) 405–413

Optics & Laser Technology

www.elsevier.com/locate/optlastec

A new data processing and calibration method for an eye-tracking device pronunciation system

Chern-Sheng Lin *, Kai-Chieh Chang, Young-Jou Jain

Department of Automatic Control Engineering, Feng Chia University, Taichung, Taiwan, ROC

Received 13 August 2001; received in revised form 26 November 2001; accepted 8 April 2002

Abstract

In this paper, a new data processing and calibration method for a pronunciation system of an eye-tracking device is described. The eye-tracking device was created using both head mounted display (HMD) technology and remote operation capabilities. A pattern recognition computer program was used to distinguish the pupil position and calculate its coordinates.

This system can be adapted to provide a digital speech function. A new method for processing the image of the eye in the PC-based system was also developed. With one video CCD camera and frame grabber analyzing a series of human pupil images while the subject is gazing at the screen, an auto-calibration algorithm is used to obtain the direction of the eye gaze in real time. The computers provide the speech sound according to the location where the eye gazes exceed 0.5 s. The availability of multipurpose in this eye-tracking system with very simple equipment will be reconfirmed for future advanced research.

Keywords: Eye-tracking device; Head mounted display; Remote operation; Auto-calibration algorithm

1. Introduction

The idea of conveying instructions and probing the impact of psychological changes through the variance in eye movement has stimulated many research studies. Because of the extensive improvements in computer speed and efficiency in recent years, the ability to track and determine pupil position has advanced tremendously. The ultimate goal of this research is to develop a human–machine interface that operates simply, easily and is accessible for allowing machine orders to be communicated by the eye.

There are many algorithms and configurations for eye-tracking devices. For example, Kaufman [1] and Gerhardt [2] in 1993 and 1994 created an eye-tracking apparatus and method employing grayscale threshold values. With an LED light source, the user's point of view on the computer screen is estimated, based on the vector distance from the center of the glint to the center of the eye in a digitized eye image. In Xie's experiment in 1994 [3], he used blue and minimum morphologic technologies for search preprocessing. It takes 3–4 s to capture the eye window using a 33 MHz CPU PC. Kirby and Sirovich [4] use the Karhunen–Loeve procedure for experiments on searching the characterization parameters of the human face. Grattan and Palmer [5] and Grattan et al. [6] developed a microcomputer-based system for the disabled that relied upon differential reflection from the eyeball. The input blink could be used to activate data presentation as a binary tree or matrix scan.

Our eye-tracking device was initially developed as a non-contact optical eye blink measuring system. It can achieve a higher rate for correct pupil position identification, by analyzing a sample eye image obtained from the CCD. This system will not cause contact injury or discomfort to the eyes. The essential core of the eyeball-tracking technology is the eye feature extraction system. In this experiment, we used a diagonal-box checker search [7]. Prior to the diagonal-box checker search, we determined the feature subset and features that were needed to save search time. The opening-and-closing action of the eyes results in other control commands that are used for driving the robots or dialing a phone system. Here we designed a head mounted display (HMD) eye-tracking system (Fig. 1). This system was coupled with a CCD and a small light source. The computer processed an eye image captured by the CCD automatically to acquire the activity of the user's eye. The program in the computer locates the center point of the pupil in these images. After transferring

* Corresponding author. Tel.: +886-4-4517250; fax: +886-4-4519951.
E-mail address: cslin@auto.fcu.edu.tw (C.-S. Lin).

0030-3992/02/$ - see front matter
PII: S0030-3992(02)00038-5

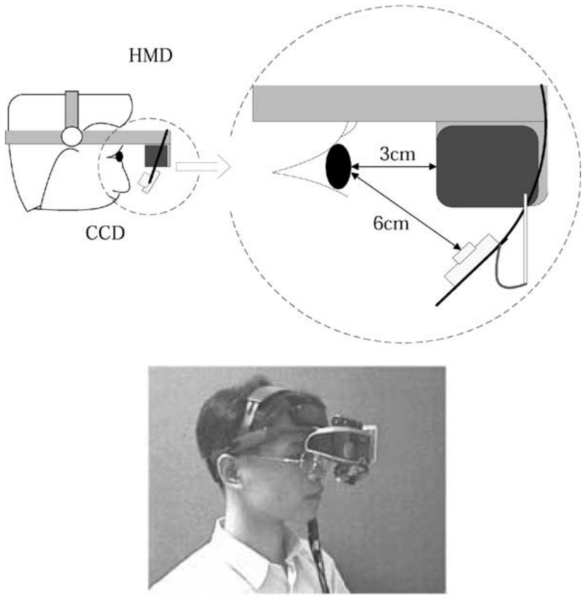


Fig. 1. HMD type eye-tracking device.

the pupil center coordinates in the images to the display coordinate, the point at which user gazes on the display is acquired. The position of the mouse cursor on the display can be controlled by eye movements. The HMD-type eye-tracking system is integrated with numerous application programs, for example, a phonetic system, Chinese and English input programs, and dial system. This system can also be programmed for remote operation (Fig. 2).

The user can communicate to others by synthesized voice, expressing by Chinese or English, and can even make a telephone call or browse the Web with his eyes. By the help of robot, the user has the ability to drag and drop objects on the computer screen using only his eyes.

This system is an improvement in accuracy and stability and is more convenient in operation. The experimental results prove that the point of gaze on the screen and the system coordinate calculations have a very linear

Fig. 3. The main diagram of the Eye-tracking system.

relationship. Compared with other eye-tracking systems, this system, plus the PC and peripheral equipment, will cost less than US$ 2,000 dollars and rate with the patients. During June of 1998–2001, many motor neuron disease patients commended our system after visiting our laboratory and observing this pupil-tracking system in operation. He considered it would be a great help to other patients with nerve-related disabilities. If this system is connected to a voice data bank its affiliated application functions and system operational processes can be simplified even more. This suggestion has deepened our confidence in this research. This system can be of great assistance to many aged, sick and disabled persons who cannot clearly express themselves.

2. Statistical data processing for decisions

In this technical process, the most important procedure is the establishment of a reliable search method. How to correctly locate the pupil position and gaze direction of the eye is the most important issue in our research on eye-position tracking. Therefore, we adopted the following search methods to execute this task.

Fig. 3 shows the primary eye-tracking system diagram. When the image is being processed, it is more convenientto proceed with the object separated from the unnecessary

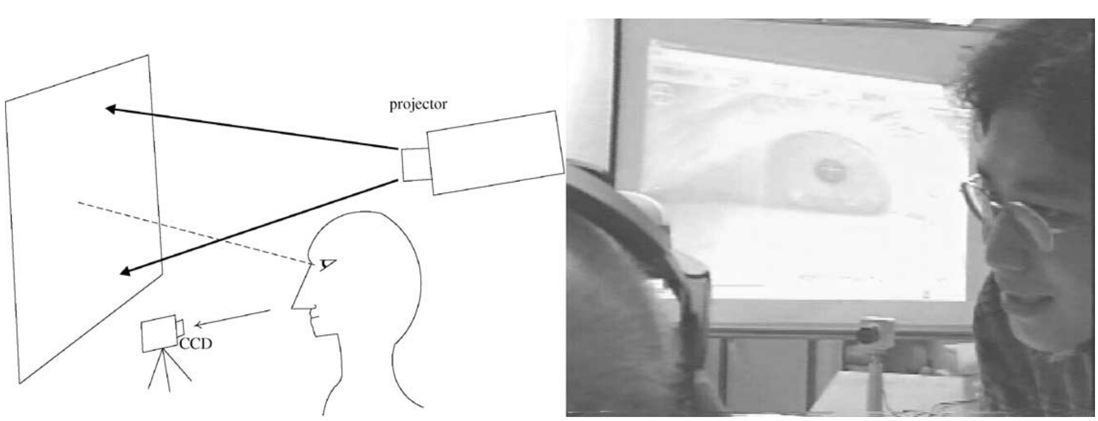


Fig. 2. Remote type eye tracking device.

(a) Upper

(b) Left

(c) Middle

(d) Right

(e) Down

Fig. 4. (a)–(e) The five calibration reference points and the pupil position.

and disruptive background. The pupil image was treated using a threshold process. The red portion is the pixel point with a gray value lower than an established threshold value, i.e., the darker part of the eye image. Next, the mean value from the pixel dimensions is calculated to acquire the central pupil value. This will facilitate the next step of searching to confirm the moving point.

In the pupil position calibration step, the statistical and vague assemblage concepts are adopted to correctly judge the gazing position of the pupil. The examples in Fig. 4, are the series diagrams of the eye looking at the calibration point. The location of the pupil is calculated, compared with the previous point and mapped to the point of screen.

Fig. 4 (a)–(e) shows the five calibration reference points relative to the pupil position. In the first stage of pupil position calibration, the coordinate data of x-direction is implemented. Initially, the average dimension (x_c, y_c) when the user looks at the central point 2 of the diagram is taken. The data (x_c, y_c) is regarded as the most exact data for the advanced calibration. Next, the estimated watching dimension (x_m, y_m) is calculated when the user looks at the point neighbor to point 2. The data (x_m, y_m) is an ideal value and depend on the resolution of the system. The distance

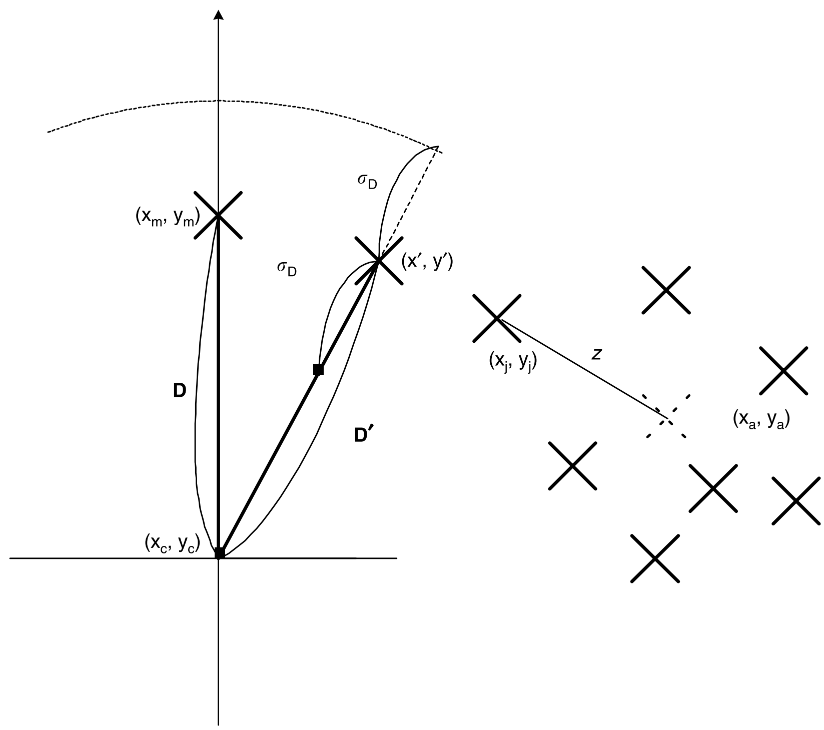


Fig. 5. The geometric relation of the gazing points.

between point (x_c, y_c) and (x_m, y_m) is calculated to be D. This value can be acquired from the input gazing points (x_i, y_i) and the average value (x', y') of them. The distance between point (x', y') and the central point is D' (Fig. 5). In the input judge, a standardized difference treatment is taken toward the x-axis dimension in order to set up a reasonable area, using the following formula:

$$\sigma = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(x_i - x')^2}. \quad (1)$$

Thus, the sampled range can be taken as follows:

$$x' - \sigma \leqslant x_i \leqslant x' + \sigma. \quad (2)$$

The above step will be used to eliminate the out of range data caused by the unconsciously gazing position in pupil movement.

From the sampling data and the central distance value D', we can obtain a standardized difference of the distance σ_D, and at the same time, to take σ_D to be the limited span, in order to ensure that the user concentrates within this range. The Λ-function diagram method, as introduced in the modified fuzzy combination theory, is then used to calculate the corresponding D' diagrams, as illustrated below

$$U(D',z) = \begin{cases} 0, & \text{when } z \geqslant D' + \sigma_D, \\ \dfrac{(D' + \sigma_D) - z}{\sigma_D}, & \text{when } 0 \leqslant z < D' + \sigma_D, \end{cases} \quad (3)$$

Table 1
The relationship between the gazing coefficient and range of $U(D',z)$

Range of $U(D',z)$	Coefficients k	Coefficients kn
$U(D',z) > 1$	$U(D',z)$	0
$1 > U(D',z) > 0.5$	$U(D',z)$	0.1
$0.2 < U(D',z) < 0.5$	$U(D',z)$	0.2
$U(D',z) \leqslant 0.2$	0	0.5
$U(D',z) \leqslant 0.2$	0	1

where $D' + \sigma_D$ is the distance between the proofreading and central points. Notation z is the corresponding distance between the new gazing point (x_z, y_z) and the central point. Two gazing coefficients k and kn are produced, according to the size of the sampled point. This system adopts the principle as follows:

According to the judgement methods in Table 1, the sample amounts can be determined using different threshold values. With these kinds of parallel judging methods, the dimension of the pupil position can be calculated more accurately. If there are many consecutive instances in which the input value of the gazing coefficient kn is 1, we can judge that the user's eye has left the gazing point. The previous accumulated times are then canceled and the process begins again. This ensures that the user actually gazed at the proofreading point.

Let $K_1 = \Sigma kn$, $K_2 = \Sigma k$.

The judging procedures are:

If $K_1 > T_1$, calculate the proofreading point, reset K_1, K_2.

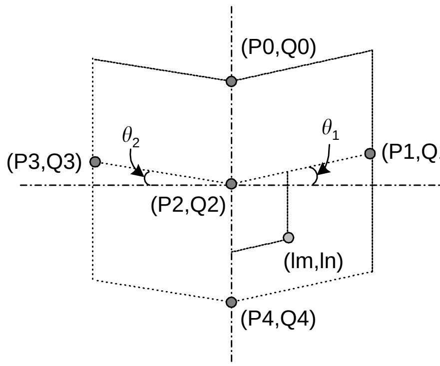


Fig. 6. The coordinate transform by linking the five calibration points as principle axis.

If $K_2 > T_2$, the user's eye has left the gazing point, reset K_1, K_2.

Assume $K_1 > T_1$ when the sample number is n. Because the data is blurred, a method must be adopted that can solve the blur. This method makes use of the center of the picture area and the weighting value of the projection to the axis. Then a corresponding value (x_a, y_a) for the gazing point in the eye image can be produced. The equation is as follows:

$$x_a = \frac{\sum_j^n x_j \times U(D', z_j)}{\sum_j^n U(D', z_j)}, \tag{4}$$

$$y_a = \frac{\sum_j^n y_j \times U(D', z_j)}{\sum_j^n U(D', z_j)}. \tag{5}$$

x_j, y_j is the coordinate value in the limited range of the jth sample. The internal range of the denominator is very close to the area of the gazing coefficient picture.

We can adopt this value (x_a, y_a) to the temporal buffer. Then we can get the necessary value in real gazing point of the practical screen in the following method.

3. Geometric operation method

The ERICA system was calibrated from the relative positions of the glint and pupil center (lookpoint vectors) for nine points of view [8]. Here, five points in the calibration process were adopted. After a standard calibration, we can obtain the reference locations (P_0, Q_0), (P_1, Q_1), (P_2, Q_2), (P_3, Q_3), (P_4, Q_4) for the center of the eye corresponding to gazing points 0, 1, 2, 3 and 4 in Fig. 4. Fig. 6 shows the coordinate transformation using five calibration points as the endpoints and central point of principal axis.

The relationship between the input data, pupil center (m, n), and five calibration reference points are based upon the following assumption:

$$p_m = m - P_2, \quad p_n = n - Q_2, \tag{6}$$

if $p_m \geqslant 0$, then

$$l_m = p_m / \cos\theta_1, \quad l_n = p_n + p_m \times \tan\theta_1, \tag{7}$$

if $p_m < 0$, then

$$l_m = p_m / \cos\theta_2, \quad l_n = p_n - p_m \times \tan\theta_2, \tag{8}$$

where

$$\theta_1 = a\sin\left(\frac{Q_2 - Q_1}{\sqrt{(P_2 - P_1)^2 + (Q_2 - Q_1)^2}}\right), \quad \theta_2 = a\sin\left(\frac{Q_2 - Q_3}{\sqrt{(P_2 - P_3)^2 + (Q_2 - Q_3)^2}}\right), \tag{9}$$

where l_m and l_n are the coordinate values after transformation.

Set point (P_2, Q_2) as the new original point. If the pupil center (m, n) is at the right side of (P_2, Q_2), then the direction (P_2, Q_2) to (P_1, Q_1) is the relative x-axis. If the center of pupil (m, n) is at the left side of (P_2, Q_2), then the direction (P_2, Q_2) to (P_3, Q_3) is the relative x-axis. The last equations to obtain the gazing point at the screen (D_m, D_n) can be written as follows:

1. if $p_m \geqslant 0$ and $l_n \geqslant 0$, then

$$D_m = X_1 - \frac{l_m}{\sqrt{(P_2 - P_1)^2 + (Q_2 - Q_1)^2}}(X_1 - X_0)C_{11},$$

$$D_n = Y_1 + \frac{l_n}{Q_4 - Q_2}(Y_2 - Y_1)C_{12}. \tag{10}$$

2. if $p_m \geqslant 0$ and $l_n < 0$, then

$$D_m = X_1 - \frac{l_m}{\sqrt{(P_2 - P_1)^2 + (Q_2 - Q_1)^2}}(X_1 - X_0)C_{21},$$

$$D_n = Y_1 + \frac{l_n}{Q_2 - Q_0}(Y_1 - Y_0)C_{22}. \tag{11}$$

3. if $p_m < 0$ and $l_n \geqslant 0$, then

$$D_m = X_1 - \frac{l_m}{\sqrt{(P_2 - P_3)^2 + (Q_2 - Q_3)^2}}(X_2 - X_1)C_{31},$$

$$D_n = Y_1 + \frac{l_n}{Q_4 - Q_2}(Y_2 - Y_1)C_{32}. \tag{12}$$

4. if $p_m < 0$ and $l_n < 0$, then

$$D_m = X_1 - \frac{l_m}{\sqrt{(P_2 - P_3)^2 + (Q_2 - Q_3)^2}}(X_1 - X_0)C_{41},$$

$$D_n = Y_1 + \frac{l_n}{Q_2 - Q_0}(Y_1 - Y_0)C_{42}, \tag{13}$$

where C_{11}, C_{12}, C_{21}, C_{22}, C_{31}, C_{32}, C_{41}, C_{42} are the system coefficients.

Table 2
Corresponding horizontal coordinate D_m or vertical coordinate D_n

Horizontal	1	2	3	4	5	6	7	8	9	10	11	12	13
D_m	75	112	153	195	241	293	332	376	425	468	512	561	599
Vertical	1	2	3	4	5	6	7	8	9	10	11	12	13
D_n	60	81	99	119	145	171	184	211	244	279	302	326	349

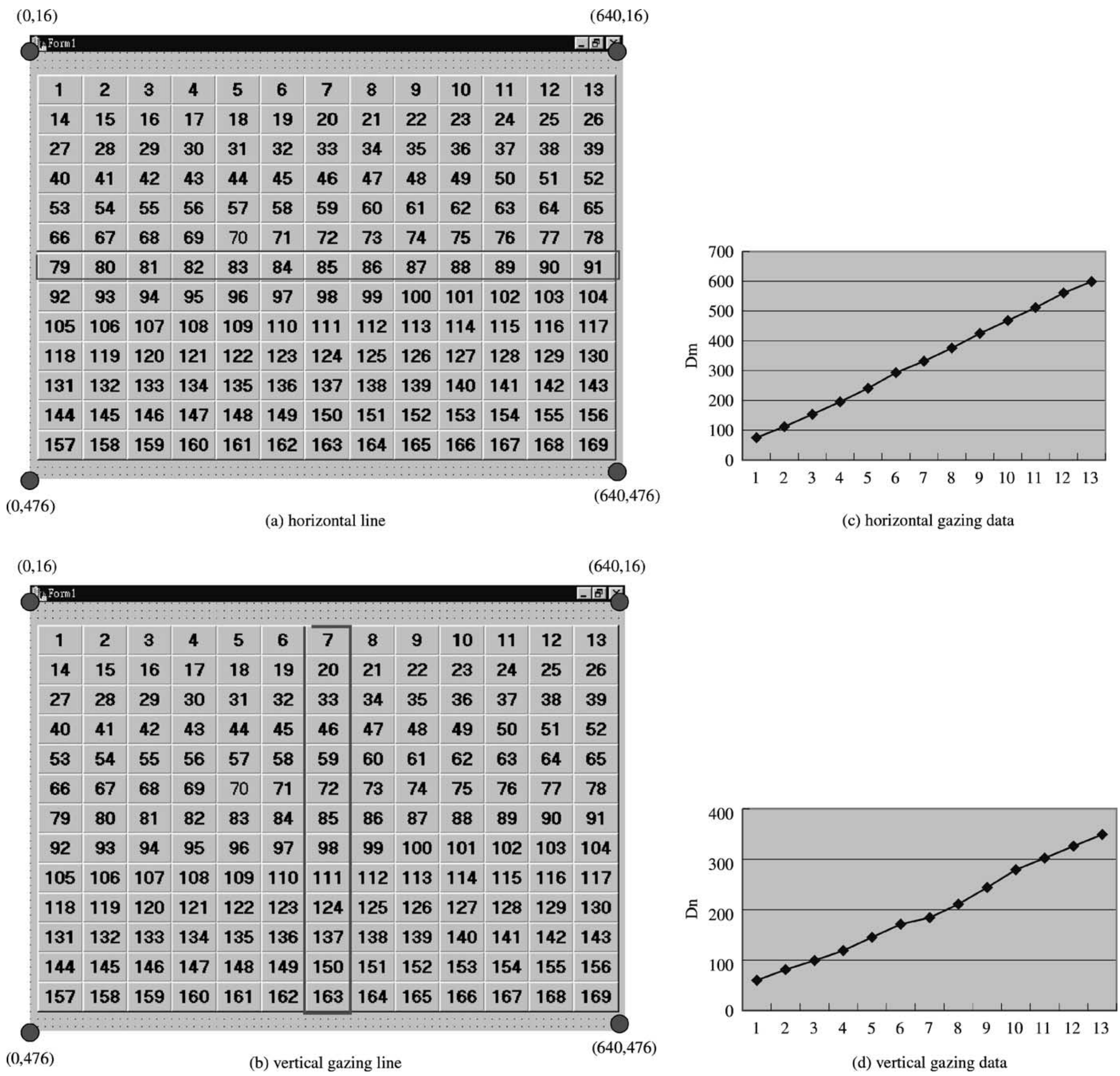


Fig. 7. The linearity of the horizontal and vertical gazing line.

Factor R can be calculated to evaluate the linearity of this system

$$R = \frac{\sum_{i=1}^{n} [(x_i - \bar{x})/S_x][(y_i - \bar{y})/S_y]}{n-1}, \tag{14}$$

where $\bar{x}$ and $\bar{y}$ are the average value of the input data. S_x and S_y are the standard deviations. R will be close to 1 when the linearity of the system is good.

Fig. 8. The Language-selection of the pronunciation system.

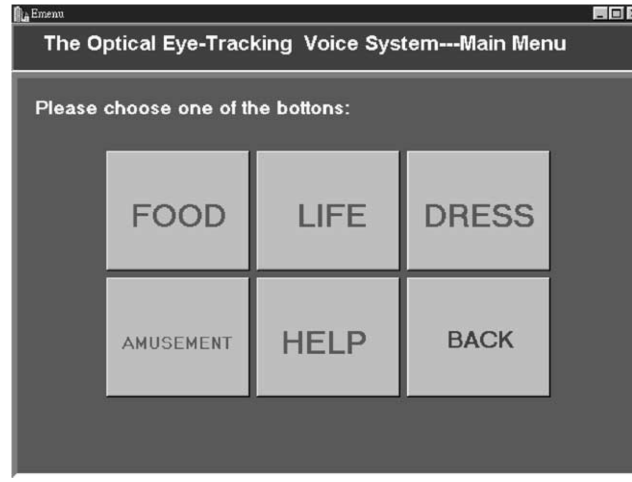


Fig. 9. The category-selection of the pronunciation system (a) Drink (b) Dress (c) Life (d) Amusement (e) Help.

4. Experimental methods and equipment

The hardware used for the eye-tracking device was composed of separable components. When the user puts on the HMD, the CCD installed under the HMD transmits the eye image to the frame grabber. Therefore, the image of the users' eyes can be seen on the HMD screen. The computer operates in real time to calculate the pupil's center position and movement speed. In the remote eye-tracking system, the user does not have to put on any instruments. It is only necessary to immobilize the head, which is a much easier method for a severely handicapped patient. The CCD camera at the front will acquire the eye image and send this image through to the LCD projector screen in front of the user. Greater space will be required for the screen installation.

Optical aberrations cause a straight eye movement to be imaged as a curved eye movement. This new geometric operation can rectify such an image into a rectangular coordinate system. We tested this system under worst case conditions, by observing 13 points on a horizontal or a vertical line and obtaining the corresponding D_m horizontal or D_n vertical coordinates, as shown in Table 2. From Fig. 7, the linearity is acceptable. The R factor is 0.999763 for horizontal linearity and 0.997148 for vertical linearity.

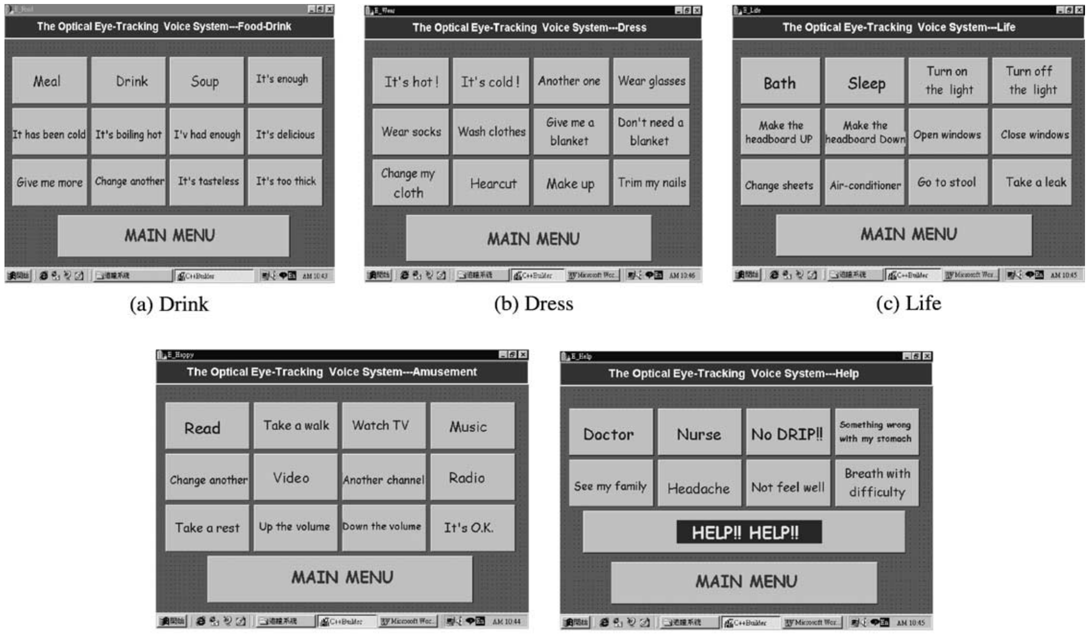
In Fig. 3, the white frame represents the search boundary. The system will auto-adjust the position of the white frame to make the pupil diagram drop onto the tableau inside the white frame. A thresholding processing is applied to the image inside the white frame in order to locate the center of the pupil. There are two buttons in the upper right corner. When the diagram is started, it will automatically "Focus" on the "start". After pressing the "ENTER" key, the pupil-tracking system will begin analyzing what will appear sequentially in the upper right, left, middle and right as well as the lower corner in the screen, as shown in Fig. 4.

After the pupil movement calibration is completed, the pupil position tracking screen is minimized and the voice system screen is activated. During this transition, if one closes his eye for 10 s or selects the "ESC" key, the pupil-tracking system screen will reappear. This allows the user to repeat the pupil-position tracking judgement procedures.

After calibration, the pronunciation system is activated. The opening diagram of the pronunciation system provides the user with language choices (there are currently two choices: Chinese or English). A diagram of this system is shown in Fig. 8. After choosing the language, the system enters the classified-choice diagram. The user can select the following screens: diet, dialing, etc. Each item is connected to an independent vision screen, as Fig. 9 shows. After selecting one of these items, there are 12 common wordings (Fig. 10) for viewing. After the user chooses a specific button, that item will begin the pronunciation system. The computers provide the speech sound according to the location where the eye gazes exceed 0.5 s.

Under the category-selection diagram, there is a function for phone dialing. After opening, a simulated telephone will appear and transmit the information by eye gaze, as shown in Fig. 11. The patient can seek help through the system during urgent situations. In order to enable the user to write sentences they wish to express, the system provides many icons in Chinese or English characters, as shown in Fig. 12. In the English alphabet board, the 26 letters were divided into two pages and the vowels are placed at the top of the screen. It is much easier to select alphabet characters than Chinese. After selecting the alphabet, some words will appear forming a complete sentence. This function provides great help to patients for expressing themselves.

At present, the majority of the patients employed in this study were "Motor Neuron Disease (ALS)" patients. Because these kinds of patients are palsied and unable to speak, their expectations are that this system will provide them with a tool for communicating with the external world. At Taipei



(a) Drink (b) Dress (c) Life

(d) Amusement (e) help

Fig. 10. (a)–(e) The independent version screen of the pronunciation system.

Fig. 11. The phone dialing diagram of the pronunciation system.

Fig. 12. The diagram of the English spelling.

Veterans General Hospital, this system has been used in a clinical test on patients, as shown in Figs. 13 and 14. Initially, the results were not as expected because the gazing ability of the patients was not acute. However, we have obtained a great deal of additional information regarding visual acuity for use with this system.

After the clinical test, we found that the contrast in the patient's eye image was not very good, their gazing ability was poor and that they were easily distracted. The extension and reaction speed of their pupils were low, as shown in Fig. 14. In order to compare the results of different users on the HMD and remote eye-tracking device to operate the pronunciation system, the operators were required to voice "Help!!" in the "Help" window. The time from visually acquiring the word "Help" to successful acquisition was recorded. This operation required the patient to gaze at different points on the screen 10 times correctly. After the user's first trial, a 10-min practice was given. The test was then conducted and recorded again. The statistical results are listed in Table 3.

From the statistics, an able-bodied youthful user can operate this system quickly after practice. It is also not difficult for nearsighted people to operate this system if they wear their glasses. In the practical test of the pronunciation system for the eye-tracking device, the correct rate of eye selection is quite high. The data base of this voice system

Table 3
The statistics for patient acquisition time on the eye-tracking device to choose "Help!!" in the "Help" vision screen

TYPE	HMD		REMOTE	
USER	Before practice (s)	After practice (s)	Before practice (s)	After practice
Normal people ($<$ 45 years old)	35	19	46	24
Normal people (more than 45 years old)	43	25	51	32
Nearsighted people (wear glasses)	34	21	42	23
Patient	$>$ 600	510	$>$ 600	560

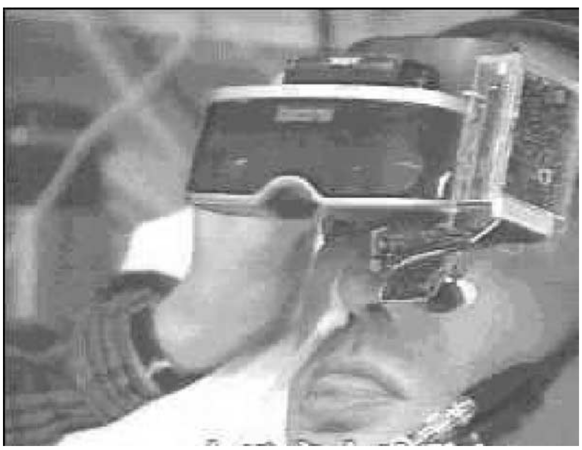


Fig. 13. The operation of the HMD type eye-tracking system.

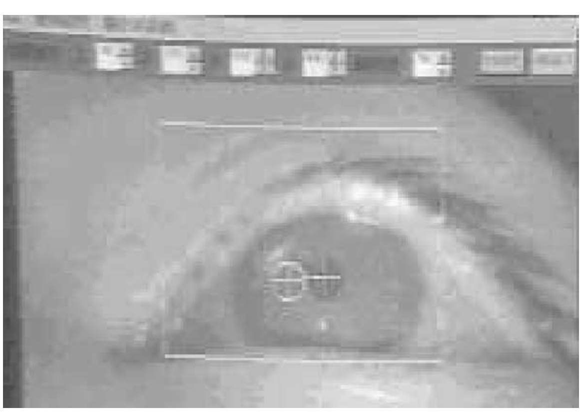


Fig. 14. The pupil image of the patient.

can be revised according to the user's need. In the future, we expect to decrease the time of the eye calibration operation and simplify the search procedure for the pupil position tracking system. For the voice system, the selected items and functions in the screen will be increased along with the events in each voice database in order to enable this system to become more functional.

5. Conclusion

When this system is connected to a non-contact eye-tracking device and operated under soft light, the location of pupil can be obtained in real time. There is no difference between this and operator reading under a standard reading lamp. For this reason, the patient operators expressed no concern about eye strain. Currently, some ALS patients have begin using this eye-tracking system with the help of the Taiwan Motor Neuron Disease Association. Additionally, tests are being conducted with other types of patients on this system. We expect to be able to design a system that will enable the disabled to operate more easily and open a larger window for their future.

Acknowledgements

This work is sponsored by the National Science Council, Taiwan, Republic of China under Grant no. NSC 88-2623-D-035-003.

References

[1] Kaufman. Apparatus and method for eye tracking interface, US Patent 5360971, 1993.
[2] Gerhardt. Eye tracking apparatus and method employing grayscale threshold value, US Patent 5481622, 1994.
[3] Xie X, Sudhaker R, Zhuang H. On improving eye feature extraction using deformable templates. Pattern Recognition 1994;27(6):791–9.
[4] Kirby M, Sirovich L. Applications of the Karhunen–Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Machine Intell 1990;12(1):103–8.
[5] Grattan KT, Palmer AW. Interrupted reflection fiber optic communication device for the severely disabled. J Biomed Eng 1984;6:321–2.
[6] Grattan KT, Palmer AW, Sorrell. Communication by eye closure—a microcomputer-based system for the disabled. IEEE Trans Biomed Eng BME 1986;33:977–82.
[7] Chern-Sheng Lin, Chih-Chung Chien, Nanjou Lin, Chiao-Hsiang Chen. Search for measuring one's blink in eyeball tracking device. Opt Lasers Technol 1998;30(5):295–301.
[8] White KP, Hutchinson TE, Carley JM. Spatially dynamic calibration of an eye-tracking system. IEEE Trans Syst Man Cybern 1993;23(4):1162–8.