Eye-Gaze Detection from Monocular Camera Image Using Parametric Template Matching

Ryo Ohtera, Takahiko Horiuchi, and Shoji Tominaga

Graduate School of Science and Technology, Chiba University,
1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan
ryotera@graduate.chiba-u.jp, {horiuchi,shoji}@faculty.chiba-u.jp

Abstract. In the coming ubiquitous-computing society, an eyegaze interface will be one of the key technologies as an input device. Most of the conventional eyegaze tracking algorithms require specific light sources, equipments, devices, etc. In a previous work, the authors developed a simple eye-gaze detection system using a monocular video camera. This paper proposes a fast eye-gaze detection algorithm using the parametric template matching. In our algorithm, the iris extraction by the parametric template matching is applied to the eye-gaze detection based on physiological eyeball model. The parametric template matching can carry out an accurate sub-pixel matching by interpolating a few template images of a user's eye captured in the calibration process for personal error. So, a fast calculation can be realized with keeping the detection accuracy. We construct an eye-gaze communication interface using the proposed algorithm, and verified the performance through key typing experiments using visual keyboard on display.

Keywords: Eyegaze Detection, Parametric template matching, Eyeball model, Eyegaze Keyboard.

1 Introduction

Human eyes always chase an interesting object. Gaze determines a user's current line of sight or point of fixation. So, the direction of the eyegaze can express the interests of the user, and the gaze may be used to interpret the user's intention for non-command interactions. Incorporating eye movements into the interaction between humans and computers provides numerous potential benefits. Moreover, the eyegaze communication interface is very important for not only users in normal health but also severely handicapped users such as quadriplegic users. Although the keyboard and the mouse are used as an interface of the computer, it is necessary for us to move the input device with the hand. Therefore, the substituted input device is necessary for the person who owed the handicap.

The eyegaze detection has progressed through measurements of infrared irradiation and myoelectric potential around eyes. Gips et al. proposed a detection algorithm based on EOG method [1] in which the myoelectric potential following the motion of eyeball can be measured by electrode on the face. However, the influence of the electric noise

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 708–717, 2007.

embedded in miniature potential is not negligible. Moreover, specific instruments are required for measurements, and the load for the user is also not negligible. Cornea-reflex-based systems were also proposed in Refs.[2]-[4]. Those systems require that the environment illumination must suppress the extraneous reflection. T.N.Cornsweet and H.D.Crane presented a simple and high-precision eyegaze detection system with the jaw board and the headrest. In this detection, the motion of head can be cancelled by utilizing 1st and 4th Purkinje images from affecting inversely each other on the motion of head [5]. However, the method using these cornea reflection images had to prepare a specific optical device such as the infrared irradiation devices to take picture of the iris image as a high luminance area or a low brightness area.

Recently video-based techniques have been studied. Compared with non-video-based gaze tracking methods as mentioned above, video-based gaze tracking methods have the advantage of being unobtrusiveness and comfortable during the process of the gaze detection. Kawato et al. presented a gaze detection method using four reference points which put on the face and three calibration images [6]. Matsumoto et al. presented a real-time stereo vision system to estimate the gaze direction [7]. Wang et al. presented a method for estimating the eyegaze by measuring the change of the contour of the iris [8]. Generally, the big-eye approaches based on the precise measurements of the eye are expensive [8],[9], because they require a pan-tilt/zoom-in stereo-camera with sufficient resolution to measure accurately iris contour or pupil. The methods using the movement of the iris captured by a low resolution single camera under the condition of fixing head pose are described in Refs.[10],[11]. These methods used a luminance gradient for extract the iris semicircle and eyes corners. Therefore, it is necessary to specify rough position of eyes beforehand. The iris extraction is very important in the algorithm that detects the gaze from the movement of the iris. The authors proposed a precise iris extraction algorithm based on the Hough transform [12]. However, the method cannot detect movement in sub-pixel and the Hough transform requires a heavy operation time.

In this study, we concentrate on a video-based approach, and develop a simple eyegaze detection system using only a monocular video camera. The proposed method does not require any specific equipment excluding the monocular video camera, and has psychologically a light burden for a user. In our algorithm, the rotation model for eyeball is constructed through traditional physiological models which are Emsley's eyeball model [13] and Gullstrand's model No.2 [14]. In the system, the eyegaze angle to optical axis is calculated by using the amount of the movement at the center of the iris after the calibration. Then, a coarse-to-fine parametric template matching method [15] is performed to extract the user's iris in robustness, high speed, and sub-pixel accuracy.

The remaining sections are organized as follows: In Sec.2, a rotation model for the eyeball is defined, and an eyegaze estimation algorithm is described. In Sec.3, an iris detection algorithm which is the key technology for estimating the eyegaze is proposed. In Sec.4, the experimental results show the performance of the proposed gaze detection system. In order to verify the performance, in Sec. 5, the proposed gaze detection algorithm is applied to an eyegaze communication interface.

2 Gaze Estimation Model

2.1 Gullstrand's Schematic Eye No.2 and Emsley's Reduced Schematic Eye

In order to estimate the eyegaze, we begin to consider physiological eyeball models. Schematic eyes are standard model in the eye optics of which the parameters are provided by observed values or its approximated values for the optical parameters in dioptrics. Several schematic eyes have been proposed. Examples include LeGrand's schematic eye, Donders' reduced eye, Lawrence's reduced eye and Listing's reduced eye. In this paper, we use Gullstrand's No.2 schematic eye and Emsley's schematic eye, which can simply express the size of eyeball. Gullstrand's model, which consists of the precise model (No.1) and the non-precise one (No.2), and Emeley's model are well-known eyeball models. Gullstrand (No.2)-Emsley's reduced eye consists of one-surface cornea, two-surface lens, spherical, rotationally symmetric surfaces. Values for the accommodation-stop and the super-accommodation of eyes can be presented.

2.2 Gaze Detection Algorithm

The eyegaze can be defined as a vector directed from the center of an iris to the center of the eyeball. In this study, the amount of the movement from the center of the iris is detected, and the eyegaze angle is calculated by using the rotation model based on the Gullstrand's reduced schematic eye (No.2) and the Emsley's reduced schematic eye. Figure 1 shows the proposed rotation model. The center of the eyeball lies at the distance 13mm behind the cornea. The gonioscope width, which is the distance from the cornea to the iris, is set as 3.4mm on averaging the value from non-controlled to strong controlled eyes (3.2-3.6mm).

The length from the cornea to the eye ground is provided by 23.9mm with the Emsley's reduced schematic eye. Therefore, the length from the eyeball rotation center to the bottom of the iris becomes 9.6mm. The amount of the eyeball rotation is defined as 0 deg for gazing at the reference point, and the amount of the movement at the center of the iris calculated in Sec.3 is defined as B. For assuming that the size of a camera is very larger than that of the eyeball, the eyegaze angle θ to the front baseline is easily given by

$$\theta = \sin^{-1} B/S \qquad (1)$$

where the eyegaze detection for oblique one is done by calculating vertical and horizontal movement B of the iris, respectively, and applying to the model separately. The symbol S in Fig.1 shows the personal size of the eyeball depending on the user, and it will be calibrated in Sec.4.2. After calculating B for vertical and horizontal direction, the oblique eyegaze is detected by applying them into the models, respectively. However, there is a necessity for the conversion to apply to the model because the amount of the movement at the center of the iris is not a value of the observational measurement. When a user fixates on the reference point, the iris becomes a circle because the eyeball does not rotate. Moreover, the diameter of the iris is assumed to be 11.5mm from the model eyes. Therefore, conversion from the obtained image to the observational measurement value assumed the diameter of the cornea to be 11.5mm from the model eyes, and calculated measurement value (mm) that corresponded to one pixel. Since we have not considered the time-sequential processing between frames, the proposed process is performed for each frame.

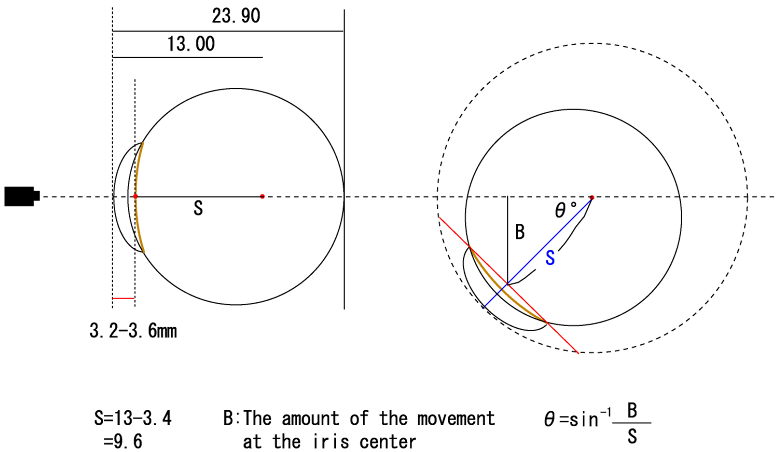

Fig. 1. The eye rotation model used in our algorithm

3 Iris Detection Algorithm

3.1 Overview of the Parametric Template Matching

As described in Sec.2.2, the eyegaze can be detected by extracting the center of the iris. So, the precise extraction of the center of the iris becomes important for accurate eyegaze detection. In the iris extraction, one of the biggest problems is how to consider the change of shape depending on the rotation of the eyeballs. In addition, high speed processing and the accuracy of the iris detection are required for real-time processing. In this paper, we use the parametric template matching method [15] for the iris extraction. The parametric template space is defined as a template space expressed by a linear interpolation of two or more given templates called "vertex templates", and high speed and high accuracy matching can be realized by the coarse-to-fine matching. The vertex templates are described in Sec.3.2 in detail.

The Zero-mean Normalized Cross- Correlation is calculated between a constructed parametric template and an object image as follows:

$$\rho(x,y) \equiv \frac{\sum_{(k,s)\in T}\left(\Delta t(x,y,k,s)\right)\times\left(\Delta g(k,s)\right)}{\sqrt{\sum_{(k,s)\in T}\left(\Delta t(x,y,k,s)\right)^2}\sqrt{\sum_{(k,s)\in T}\left(\Delta g(k,s)\right)^2}} \tag{2}$$

$\Delta t(x,y,k,s) \equiv t(x+k, y+s) - \bar{t}$

$\Delta g(k,s) \equiv g(k,s) - \bar{g}$

$T \equiv \{(k,s) | 1 \le k \le l_x, 1 \le s \le l_y\}$; $\quad l_x, l_y$: template size

Where, t, g represent the luminance value within the template image and the object image, respectively. In this paper, an iris image at gazing center position is used as the object image. Symbols $\bar{t}$ and $\bar{g}$ are mean values of each images. The correlation is calculated for all positions within the input image, and the position (x^*,y^*) with the maximum correlation $\rho(x^*, y^*)$ is detected as the iris position. Here, we use the coarse-to-fine searching algorithm for reducing processing time which is explained in the next subsection.

3.2 Construction of the Parametric Template

In this section, we define the vertex templates and construct the parametric template. We use the following seven images as the vertex templates.

(a) Three local images within an input image
(b) Four iris images when each display corner is gazed.

Images (a) are used as vertex templates for detecting the iris movement. As shown in Fig.2, three vertex templates $\hat{t}_1, \hat{t}_2, \hat{t}_3$ are local images at the position (x, y), $(x+\Delta x, y)$ and $(x, y+\Delta y)$ in the input image, respectively. Here, Δx and Δy are the sampling intervals. Images (b) are used as vertex template for robustness of geometric transformation of the iris because the shape is transformed into the ellipse by rotating

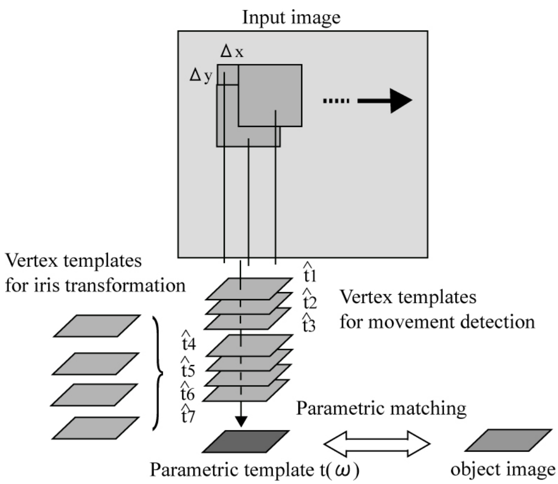

Fig. 2. Concept of the parametric template matching

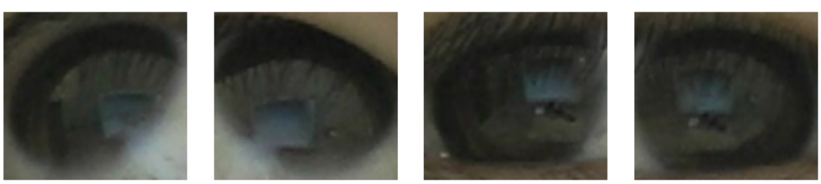

Fig. 3. Vertex template images for iris transformation

the eyeball. In Fig.2, the vertex templates are expressed as $\hat{t}_4, \hat{t}_5, \hat{t}_6$ and $\hat{t}_7$. Figure 3 shows an example of a set of vertex images for iris transformation. When the rotation angle is large, the iris is concealed by the eyelid. In general multi template matching, all transformed iris images must be prepared beforehand. In this study, we get four vertex images by gazing corner position and express variation of the iris shape by interpolating those four images continuously. By using seven vertex images, the parametric template $t(\omega)$ is constructed by linear interpolation proposed in Ref.[15].

By using the constructed parametric template, the correlation $\rho(x, y)$ in Eq.(2) is calculated for all position in the input image. The most matched position (x^*,y^*) with the highest $\rho(x^*, y^*)$ becomes a candidate position of the iris. After that, the sampling intervals Δx and Δy decrease, and more precise matching is performed around the candidate position again. These procedures are repeated by decreasing the interval until $\Delta x = \Delta y = 1$. In the case $\Delta x = \Delta y = 1$, a sub-pixel matching can be realized. Finally, the center of the template at the most matched position is detected as the center of the iris.

3.3 Adjustment for Center of Iris

The shape of the iris can be transformed into the ellipse when the eyeball is rotate. So the center of the iris projected in the two dimensional image may shift from the center of the detected template. Figure 4 shows the eyeball which we looked at from the top. The line A-B shows the iris and the line A'-B' shows a rotated iris. Algorithm in the previous section extracts the point C as the center of the iris, because the algorithm extracts it as the geometric center point between A' and B'. However, actual center point of the iris is E in Fig.4. For more accurate eyegaze detection, we have to adjust the center of the iris. We assume that the correct center of the iris is the center of the iris when the front is gazed. Let O be the center of the eyeball in Fig.4. Then A and B can be expressed as $\left(-I, -\sqrt{S^2 - I^2}\right)$ and $\left(I, -\sqrt{S^2 - I^2}\right)$, respectively. Let S be the calibrated radius of the rotation locus. Detailed calibration process is described in Sec.4.2. Then A' and B' transformed by rotating θ degrees can be expressed as follows:

$$A': \begin{bmatrix} A'_x \\ A'_y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} -I \\ -\sqrt{S^2 - I^2} \end{bmatrix} \tag{3}$$

$$B': \begin{bmatrix} B'_x \\ B'_y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} I \\ -\sqrt{S^2 - I^2} \end{bmatrix} \tag{4}$$

In Fig.4, f, which is the x-coordinate of the point D, can be calculated as

$$f = \frac{A'_x + B'_x}{2} = \sqrt{S^2 - I^2}\sin\theta \tag{5}$$

Here D is the center of the iris in the camera image. So, the actual angle can be derived as

$$\theta = \sin^{-1}\frac{f}{\sqrt{S^2 - I^2}} \tag{6}$$

Then the accurate eyegaze vector O-E is detected.

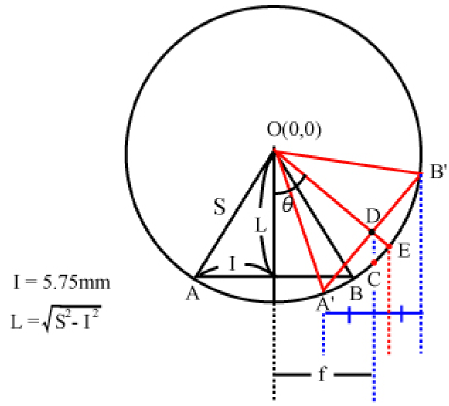

Fig. 4. Adjustment for center of iris

4 Experimental Results of Eyegaze Detection

4.1 Experimental Environment

The proposed method is here demonstrated for the display interface. In the experiment, a reference point is set on the center of display, and an observer with the naked-eyes sits in front of the reference point. The observer fixates his eye to the front, thus the effect of the direction of his face can be suppressed. Observers are three males. Each observer tested by twice. As an observer sits in front of the eyegaze monitor display, a monocular video camera mounted below the monitor observes one of the observer's eyes. The distance from the display to the eyeball of the observer sitting on the chair is set with 400mm, which is a widely usable distance. The source of light is arranged a little backward of the observer only by an overhead fluorescent lamp. The observer is irradiated from the upper behind by only the fluorescent lamp. The face image is taken with the digital video camera, Panasonic NV-GS200K(640×480), which is at the distance 120mm apart from the display. The jaw and forehead of observer is fixed on the plate.

In the experiment, 20 indices without the center index were displayed at 10 degree intervals in visual angle. The center index was used as the reference point. One index was displayed in five seconds intervals. Then, the observers gazed the displayed index. The face image was captured four seconds later after the index was displayed. The direction of the eyegaze from the reference point was calculated for each selected indices by using Eq.(1). Under the condition of fixing head pose, the iris doesn't move large. The searching region was limited from an initial position in about 2.5 times of the iris size on account of high-speed processing.

4.2 Calibration Method

In general, before the eyegaze is detected, individual calibration for each observer is performed by gazing two or more markers, here 5-20 markers, on the display. In the eyegaze detection, it is to be desired that the individual calibration is unnecessary. However, because a deviation result from some factors, the individual calibration is actually required. The guessable factors in the proposed method are as follows. Errors due to the optical system such as the position of camera

(1) Refraction at the surface of cornea
(2) Personal error due to the shape of eyeball
(3) Degree of aspheric for the surface of cornea
(4) Refracting through the grasses or contact lenses

This paper focuses on Error (3). Considering personal eyeball size in Emsley's model eye and Gullstrand's model eye No.2, we propose simple 4-point calibration using corner points of the display screen. The points are (vertical view angle[deg], horizontal view angle[deg])=(-25, 20), (-25,-20), (25,-20), (25,20). First, the display is divided into four blocks from the reference point. Next, B in Eq.(1) is calculated. Then, the parameter S in Eq(1) is personalized so that the amount of the rotation of the eyeball in the calibration process. The same adjustment procedure proceeds to the other calibration point. Then, the adjusted parameter S to four blocks is obtained.

4.3 Results of Eyegaze Detection

We compared the proposed method with a conventional algorithm in Ref.[15] under the same system condition. Table 1 shows the average error of gaze detection and processing times. The average error of the proposed method is 0.92deg for horizontal direction, 2.09deg for vertical direction. The average error of the proposed method is slightly larger than the conventional method. However, the maximum error of the proposed method is smaller. Therefore, the proposed method realizes stable eyegaze detection. Next, we compared as to the processing time. In the parametric template matching, the initial sampling intervals are set to 8-15. Then the processing times are averaged. As shown in Table 1, the processing times are drastically reduced.

The proposed method is inferior to Ref.[12] in the accuracy. However, it has an advantage in the processing time. There is a large influence in human comfort at the processing time when the eyegaze is applied to the human interface. As for the proposed method, further speed-up is expected by using information between frames.

Table 1. Average error of gaze detection and processing times

	horizontal	vertical	Processing times
Ref.[12]	0.66.deg	1.05.deg	86.3 sec
The proposed method	0.92.deg	2.09.deg	0.69 sec

5 Application for Eyegaze Keyboard

We developed an eyegaze communication interface using the proposed method. A user can operate the developed eyegaze communication system by looking at rectangular keys that are displayed on the control screen. Japanese syllabary was written on the visual keyboards. Then the simple word processing can be realized by looking at each key in turn. The size of a key is 5 deg. When the error exceeds 2.5 degrees, the detection fails. However, a user can push forward work without stress so that a letter is detected fast and correct it quickly. Figure 5 shows the eyegaze keyboard system.

Fig. 5. Eyegaze keyboard system

6 Conclusion

This paper has proposed a simple method for eyegaze detection. It is non-contact for the observer and any specific devices excluding the monocular video camera are not required. Moreover, this method requires neither the reference light nor the infrared rays light, etc. The rotation model of the eyeball was constructed. Then, we devised a simple and fast eyegaze detection algorithm by iris extraction using the parametric template matching method. In order to verify the performance of the proposed method, an eyegaze detection experiment was performed. The average of horizontal direction error was 0.92deg, and vertical one was 2.09deg. Although the accuracy was slightly bad, compared with the conventional algorithm, the processing speed was improved with about 1/125 drastically. The improvement of accuracy is future work. For more comfortable system, head-free condition is required.

References

1. Gips, J., Olivieri, C.P., Tecce, J.J.: Direct control of the computer through electrodes placed around the eyes. In: Smith, M.J., Salvendy, G. (eds.) Proc. 5th Int. Conf. on Human Computer Interaction, Orlando, FL. Published in Human-Computer Interaction: Applications and Case Studies, pp. 630–635. Elsevier, Amsterdam (1993)
2. Talmi, K., Liu, J.: Eye and gaze tracking for visually controlled interactive stereoscopic displays. Signal Processing: Image Communication 14, 799–810 (1999)
3. Hutchinson, T.E., White, K.P., Martin, W.N., Reichert, K.C., Frey, L.A.: Human-computer interaction using eyegaze input. IEEE Trans. Systems, Man & Cybernetics 19(6), 1527–1534 (1989)
4. Ohno, T., Mukawa, N., Kawato, S.: Just Blink Your Eyes: A Head-Free Gaze Tracking System. In: Int. Conf. for Human-Computer Interaction, Florida, USA, pp. 950–951 (2003)
5. Cornsweet, T.N., Crane, H.D.: Accurate two-dimensional eye tracker using first and forth Purkinje images. J. Opt. Soc. Am. 63(8), 921–928 (1973)
6. Kawato, S., Tetsutani, N.: Gaze Direction Estimation with a Single Camera Based on Four Reference Points and Three Calibration Images. In: Narayanan, P.J., Nayar, S.K., Shum, H-Y. (eds.) ACCV 2006. LNCS, vol. 3852, pp. 419–428. Springer, Heidelberg (2006)

7. Matsumoto, Y., Zelinsky, A.: An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In: Proceedings of IEEE fourth Int. Conf. on Faze and Gesture Recognition, pp. 499–505 (2000)
8. Wang, J., Sung, E.: Gaze determination via images of irises. Image and Vision Computing 19(12), 891–911 (2001)
9. Kim, K.-N., Ramakrishna, R.S.: Vision-based Eyegaze Tracking for Human Computer Interface. In: IEEE Int. Conf. On Systems, Man, and Cybernetics, vol. 2, pp. 324–329 (1999)
10. Hammal, Z., Massot, C., Bedoya, G., Caplier, A.: Eyes Segmentation Applied to Gaze Direction and Vigilance Estimation. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) ICAPR 2005. LNCS, vol. 3687, pp. 236–246. Springer, Heidelberg (2005)
11. Benoit, A., Caplier, A., Bonnaud, L.: Gaze direction estimation tool based on head motion analysis or iris position estimation. In: Proc. EUSIPCO2005, Antalya, Turkey (September 2005)
12. Ohtera, R., Horiuchi, T., Kotera, H.: Eye-gaze Detection from Monocular Camera Image Based on Physiological Eyeball Models. In: IWAIT2006. Proc. International Workshop on Advanced Image Technology, pp. 639–664 (2006)
13. Emsley, H.H.: Visual Optics, 5th edn. Hatton Press Ltd, London (1952)
14. Gullstrand, A.: Appendix II.3 The optical system of the eye, von Helmholtz H, Handbuch der physiologischen Optik (1909)
15. Tanaka, K., Sano, M., Ohara, S., Okudaira, M.: A parametric template method and its application to robust matching. Proc, Computer Vision and Pattern Recognition, IEEE, 620–627 (2000)