A Gazing Point Estimation method on the Monitor by Using the Surrounding Camera

Takeshi Iwana*, Kazuhiko Yamamoto*, Kunihito Kato*, Hitoshi Hongo**

*Faculty of Engineering, Gifu University
1-1, Yanagido, Gifu, Japan 501-1193

**HOIP, Softopia Japan & JST
4-1-7 Kagano, Ogaki, Gifu, Japan 503-8569

Abstract

In this paper, we propose the Surrounding Camera System that estimates gaze position of a user who watches a monitor. The Surrounding Camera System consists of four cameras at four places surrounding the monitor. This system realizes an interface for cursor control, window selection and other operations using gaze point estimation.

The surrounding camera system allows users to move in front of the monitor. First, using four cameras, our system detects the face region using color information and estimates the distance between the user and the system using the stereo matching. Next, face direction is estimated from each detected face by using the four directional features and the linear discriminant analysis. Then, the accuracy of face direction estimation is improved by integrating the results from the four cameras. Finally, the gazing point is determined from the distance of the user and the face direction. We experimented with the gaze detection using the surrounding camera system. The experiments showed that the errors of the gazing point were less that 6 cm even though the user moved from 70 cm to 150 cm in front of the system.

1 Introduction

In recent years, pattern recognition technologies have improved dramatically, and have been applied to many systems such as entrance check by facial image, gesture recognition system, gaze interface, and so on [1]. In this paper, we describe a system which estimates the gazing point of user on a monitor. If a user's gaze point on the monitor can be estimated by the computer automatically, the mouse cursor control, window selection are possible by the movements of the face and eyes. At present, some gaze interfaces have been proposed[2,3,4]. Their system can estimate a gaze point of a user. However, the user needs to sit in front of their systems and is not allowed to approach the system or to move away.

In order to solve this problem, we propose a system that can estimate the gazing point of the user who is moving in front of the system. This system consists of four cameras installed at four places surrounding the monitor. We call this system surrounding camera system. The surrounding camera system detects the face from four cameras respectively, and estimates the position of the user using the stereo maching. The direction of each face is estimated by the four directional features and the liner discriminant analysis. To improve the accuracy of face direction, our system integrates the estimated results from the surrounding cameras. In this paper, we describe the overview of the surrounding camera system and algorithm of gazing point estimation. In addition, we discuss the experimental results using our system.

2 Surrounding Camera System

2.1 System configuration

Figure 1 shows an overview of the surrounding camera system. The surrounding camera system captures the images from the four cameras respectively. The four cameras are installed at the four places of the monitor.

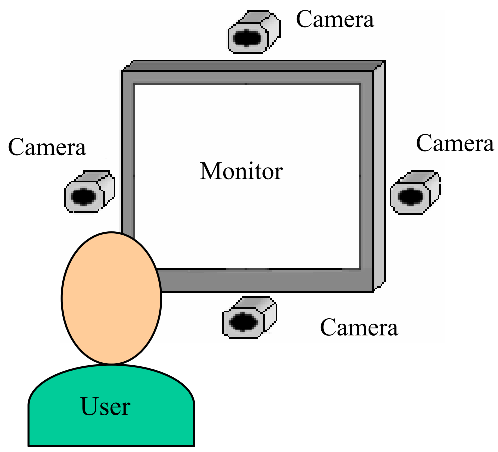


Figure 1. Overview of the system

Figure 2 shows the camera placements on this system. The four cameras are arranged as follows: Both of the distances between the upper and lower camera, and between the left and the right cameras are 59 cm. The axis of the lens of Each camera faces inside 15 degrees. All optical axis intersect at one point which is 113.9 cm from each camera, and distance from monitor is 110 cm.

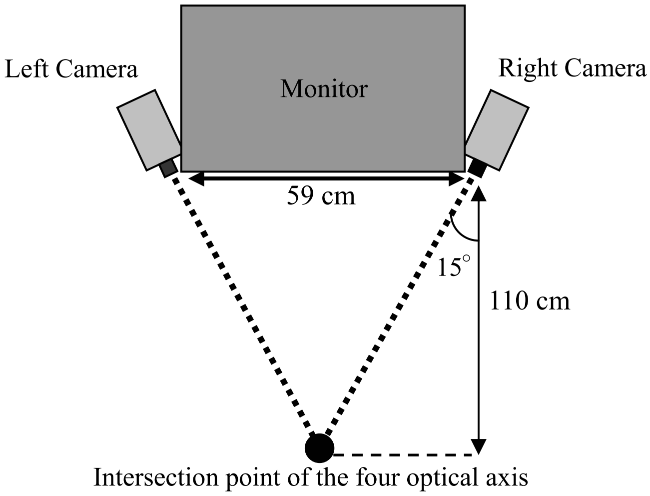


Figure 2. Camera placements

This system does not have a camera forwards to the person. Figure 3 shows an example of an input image obtained by the Surrounding Camera System. Four different viewpoints of the images are combined by the image switcher. Thus, the Surrounding Camera System can acquire the four images simultaneously. Since the results of face direction estimation are integrated from four cameras, it is possible to estimate gazing point on the monitor with high accuracy and stability.

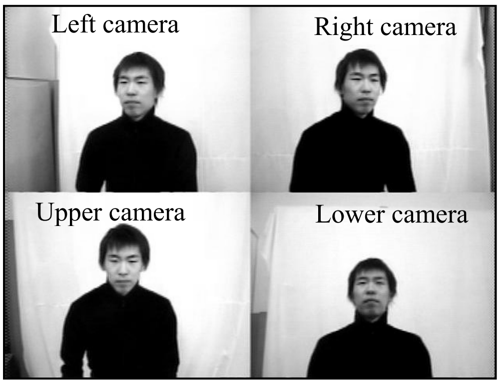


Figure 3. An example of the combined image

2.2 System flow

Figure 4 shows the system flow. First, the Surrounding Camera System detects a face from each input image. Secondly, a user position is mesured using the stereo matching and then face direction is estimated. Finally, we estimate the gazing point on the monitor by integrating face direction and user's position.

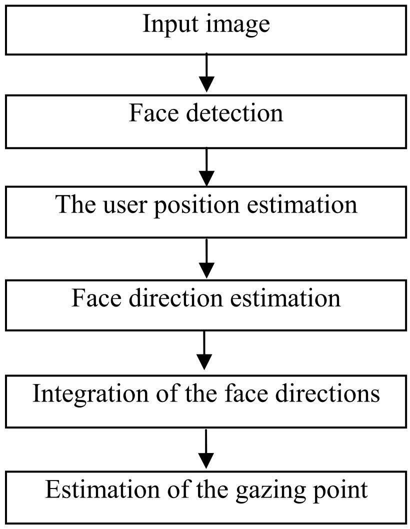


Figure 4. System flow

3 User position estimation

We have to measure a user's position, because we estimate gazing point on the monitor by integrating face direction and user's position.

3.1 Face detection

In order to detect a face for measuring the user position, we use U, V values in the CIE-LUV color system[5]. The CIE-LUV color system consists of lightness(L) and color values(U and V). Theoretically, the CIE-LUV color system is just as capable as humans when distinguishing between any two colors of the same lightness. Our method that extracts face region is as follows:

First, the range of skin UV values is decided from a sample of several subjects of asian origin. Next, we change the input image into the CIE-LUV color system from the RGB color. Our method determines the standard skin colors, which indicate the maximum number of pixels within the range of skin UV values. Figure 5 shows an example of the UV values distribution. The rectangle shows the range of skin UV values, and the position of a cross indicates the standard skin color. The distances from the standard skin color to each pixel of UV values in the input image are calculated. Figure 6 shows the input image. Figure 7 shows the UV value distance histogram from standard skin color. We extract the skin color region using the discriminant analysis. We decide that the face region is the circumscribed rectangle of the skin color region. In figure 6, frame shows skin region.

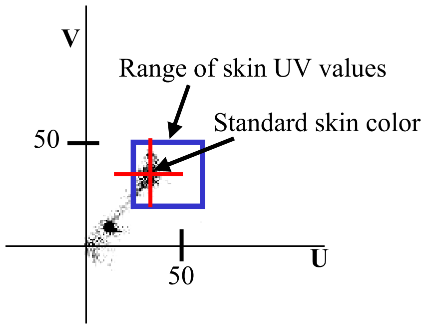


Figure 5. UV value distribution

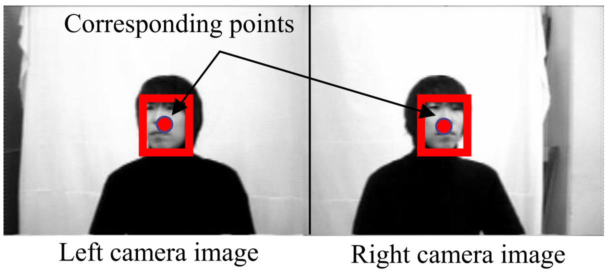


Figure 6. Face detection and the detected correspond points

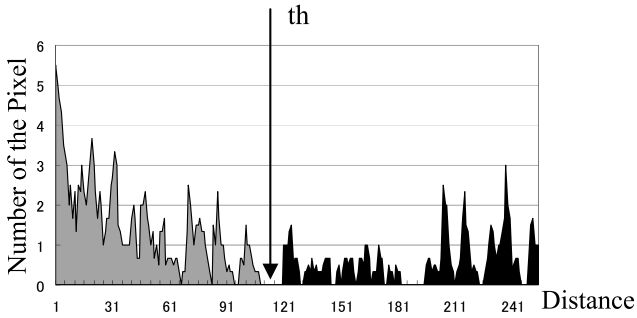


Figure 7. The UV value distance histogram from the standard skin color

3.2 Method for estimating the user's position

In this system, the user's position is measured using the stereo matching[6]. We use the camera installed on the right and left of monitor. First, we extract the face region as shown in figure 6, and we determine the center of the circumscribed rectangle of the skin color region as the corresponding point of each. Measuring the user's position *D' is* backward from actual user's position *D* as the figure 8, because the corresponding point is center of the circumscribed rectangle of the skin color region. Thus, we compensate this error by equation (1). Let user's head be a sphere, the parameter α is radius of head.

$$D = D' - \alpha \qquad (1)$$

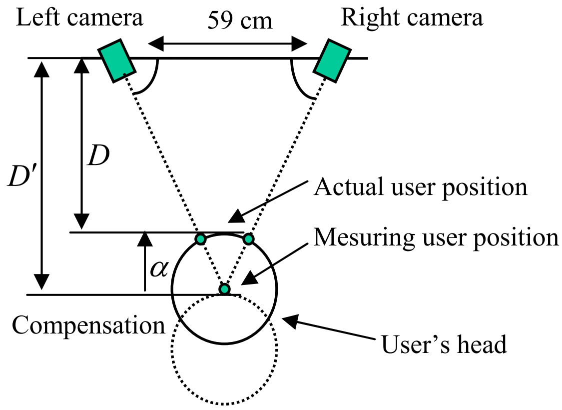


Figure 8. Measuring user's position

We experimented with measuring user's position. We collected the experimental data where the user gazed at the center of the monitor and camera installed on the upper, lower, right and left of monitor. Figure 9 shows the results of measuring user's position. Measuring the values of user's position doesn't influence the direction of the face, and user's position is measuring in the higher accuracy.

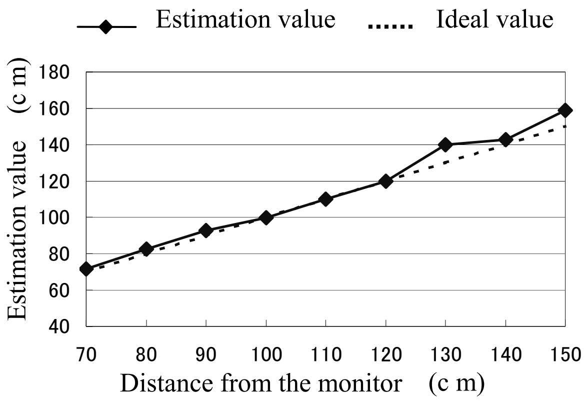


Figure 9. User's position estimation

4 Gazing point estimation

A gaze point is estimated by integrating the direction of the face and user position information.

4.1 Face direction estimation

This system distinguishes an up-down and left-right direction, 0, 15, 30 degrees of the face. Four images are observed at once. Thirteen training images are combined five different face directions. The face images for training as shown in figure 10 are acquired by extracting the skin color region from training images. The training dictionary is generated from training images, which user gaze at the center of the monitor and camera installed on the upper, lower, right and left of monitor as shown in figure 11. Thus this system can make the training dictionary without being a burden to the user.

Features of the training face images are extracted. We used the four directional feature fields[7]. Figure 12 shows the four directional feature fields. The four directional feature fields consists of four images which are divided by edge direction into horizontal, the vertical, the right top, and the left top. They are reduced to low resolution through the Gussian filter. Even if it is low resolution, this feature keeps the information of the edge directions in comparison with the feature that does not divide the direction of edge. We extract 256-dimensional (8x8x4) four directional feature fields from the training face images. Extracting features can be classified using the linear discriminant analysis [8], and we acquire training dictionary which can divide into 13 classes.

Figure 10. An example set of face images for training

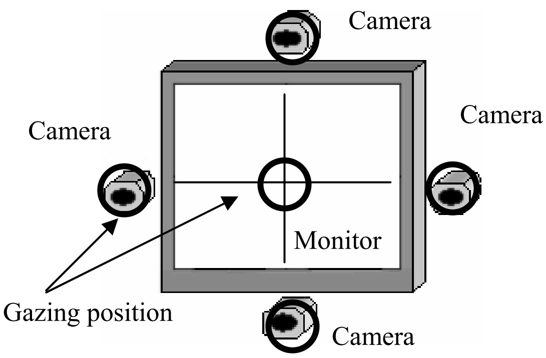


Figure 11. Generation of the training dictionary method

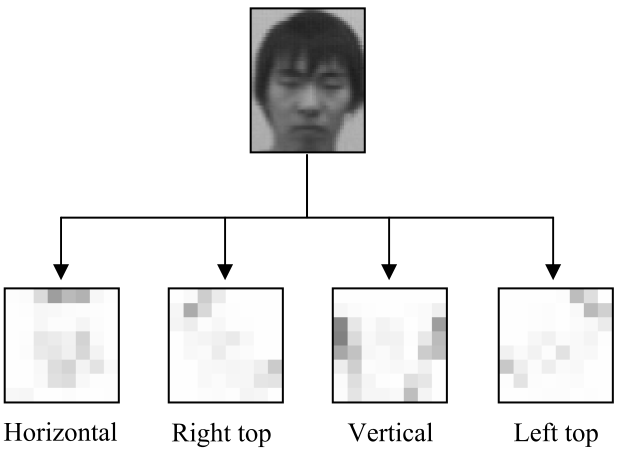


Figure 12. The four directional feature fields

We evaluated performance of the dictionary. We experimented with dividing the experimental data into 13 classes by the dictionary as shown in figure 10. The dictionary consisted of 1950 face images. We collected the experimental data which user gazed at the center of the monitor and camera installed on the upper, lower, right and left of monitor. The user position is generation point of training dictionary. The number of experimental subject was 6. The experiment result is approximately 93% success rate. we confirmed the training dictionary of this system is high accuracy.

4.2 Integration of the face direction

We extract the four directional feature fields from the input face image. Extracting features from input face image be classified by the linear discriminant analysis. We obtain coefficient matrix A and mean vector y_k from the training face images by the linear discriminant analysis. We obtain vector X_i from input face image by mapping conversion of coefficient matrix A. Thus, we calculate the distance between the vector X_i and training dictionary on the discriminant space by equation (2), and we choose higher 2 class dicθ_1 and dicθ_2 by the Euclidean distances, and then the Euclidean distances are d_1 and d_2 each. Direction of the face $\theta_{x,y}$ is calculated by angle interpolating of 2 classes. Equation (3) shows the direction estimation of the face method. The direction of the face $\theta_{x,y}$ of each camera is estimated respectively.

$$D_{ik} = \left| X_i - y_k \right|^2 \qquad (2)$$

$$\theta_{x,y} = \frac{dic\theta_1 \times d_2 + dic\theta_2 \times d_1}{d_1 + d_2} \qquad (3)$$

The direction of the face $\theta_{x,y}$ which is calculated by equation (3) is camera coodinate system as shown in figure 13(a). Thus, the direction of the face is changed into the monitor coordinate system by user's position information as shown in figure 13(b). Finally, we estimate the direction of the face by average of the face direction estimation from each camera.

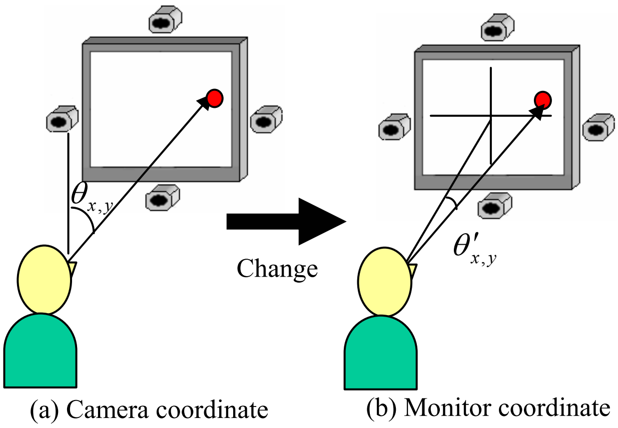


Figure 13. The direction of the face

4.3 A gazing point estimation method

We defined the direction of the gaze as the direction of the face. A gaze point $G_{x,y}$ on the monitor is estimated by integrating the direction of the face $\theta_{x,y}$ and user position information D as shown in figure 14. Equation (4) shows the gazing point estimation, and the gazing point is calculated as the origin of the center of the monitor.

$$G_{x,y} = D \times \tan\theta_{x,y} \qquad (4)$$

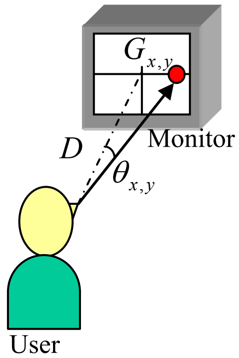


Figure 14. A gazing point estimation method

5 Experiments

5.1 Gazing point estimation experiments

We collected the experimental data which user gazed at the center of the monitor and camera installed on the upper, lower, right and left of monitor, the user gazed at each direction at three times. The user moved from 70 cm to 150 cm in front of the monitor as shown in figure 15. Number of experimental subject is 6. We only made training dictionary on fixed point which distance from monitor is 110 cm. The training dictionary is made from 1950 face images. Figure 16 shows the input image, and figure 17 shows the experimental results of gazing point estimation. In the figure 17, error means distance between the ideal the value and estimated value on the monitor.

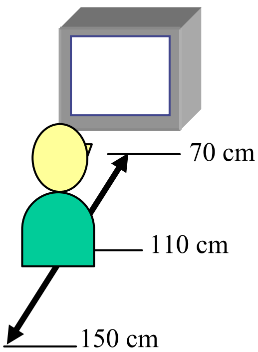
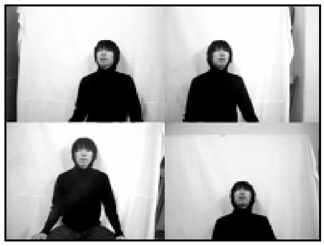


Figure 15. The user position

70 cm



110 cm 150 cm

(Left, right, lower and upper camera image in clockwise)

Figure 16. Experimental data

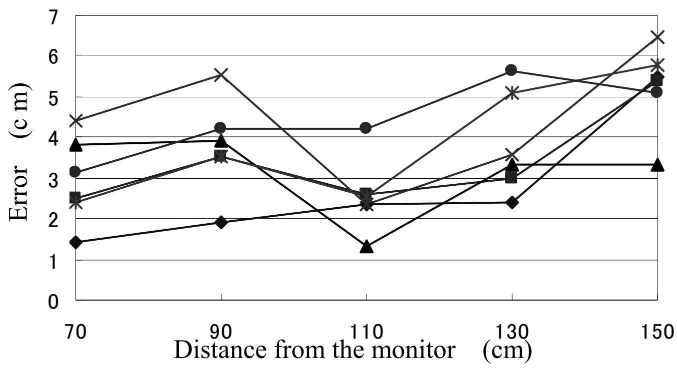


Figure 17. Experimental result

In the figure 17, gazing point estimation value becomes less accurate as the user's distance from the monitor increases. Because the estimation value of gazing point is largely influenced by the face direction estimation. The result gives as high accuracy in the distance from monitor which is the learning position of the training dictionary. This system can adapts to user position changes by the training dictionary which is made on fixed point.

5.2 Comparing the degree of freedom

The experimental data is same as section 5.1. We compared the degree of freedom of the user position by two methods. One method used the user position by measuring the distance, the other method does not use. In the method without measuring user position, user is limited to position of the training dictionary. Figure 18 shows the experimental results which we compared the degree of freedom of the user position. In the figure 18, error means distance between ideal value and estimated value on the monitor. In the figure 18, the experimental results are average values of all experimental subjects.

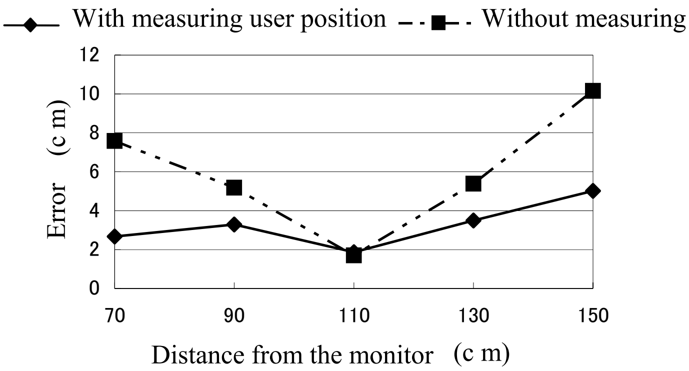


Figure 18. Experimental result

In the figure 18, estimation value of gazing point becomes less accurate as the user's distance from the learning position as 110 cm increases. The method with measuring keeps high accuracy at the position which is not learning position, but also all distances. Thus, using this system, the user can move free in front of the monitor.

6 Conclusion

In this paper, we proposed the surrounding camera system which estimates the user's gaze point in front of the monitor.

The proposed system measures use's position from the monitor, and the face direction is estimated by integrating the results from the four cameras. Using our system, we experimented with the gaze detection while the user moves in front of the monitor. Our system can estimate the gaze point with higher accuracy. This system can adapts to user position changes by the training dictionary which is made on fixed point.

As future works, we are going to improve gazing point estimation accuracy.

Acknowledgment

The authors would like to acknowledge the support of HOIP, softopia Japan, and Japanese Science and Technology Corporation (JST).

References

[1] A.Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing", IEEE Trans. on PAMI, Vol.22, No.1, pp.107-119 January 2000.

[2] Y.Matsumoto and A.Zelinsky. " An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement," Proc. FG2000, pp.499-504, 2000.

[3] T.Ohno, N.Mukawa and A.Yoshikawa, "FreeGaze: A Gaze Tracking System for Everday Gaze Interaction", Proc. Eye Tracking Research & Application (ETRA2002), pp.125-132, 2002.

[4] K.R.Park and J.Kim, "Gaze Point Detection by Computing the 3D Positions and 3D Motion of Face", IEICE Trans. INF. & SYST.,Vol.E83-D,No.4 , pp.884-894, 2000.

[5] H.Hongo and K. Yamamoto, "Method of Face and Facial Features Detection by Using Estimation of Skin Color", Proc. 5th Korea-Japan Joint Workshop on Computer Vision (FCV'99), pp.74-79, 1999.

[6] M.Ohya, H.Hongo, K.Kato, and K.Yamamoto: "Multiple Face Tracking and Zooming System", ICMI'2000, pp.168-175, 2000.

[7] K.Yamamoto, "Present State of Recognition Method on Consideration of Neighbor Points and Its Ability in Common Database",IEICE Trans. Inf. and Syst.,Vol.E79-D,No.5,pp.417-422,1996.

[8] T.Kurita, N.Otsu and T.Sato, "A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis", Proc. 11th ICPR, Vol.2, pp.213-216, 1992.