8—25

Detection and Estimation of Omni-Directional Pointing Gestures using Multiple Cameras

Hiroki Watanabe * Hitoshi Hongo * Mamoru Yasumoto *
HOIP, Research and Development Department,
Softopia Japan and JST

Kazuhiko Yamamoto †
Faculty of Engineering,
Gifu University

Abstract

We propose a multi-camera system that can detect omni-directional pointing gestures and estimate the direction of pointing. In general, when a human points at something, their target exists directly in front of the direction they are facing. Therefore, we regard the direction of pointing as the direction represented by the straight line that connects the face position with the hand position. First, the multiple cameras detect the face region by skin colors and estimate the face direction with the discrete face direction feature classes. Second, we estimate the precise direction that the subject is facing with the integrated information from multiple cameras and decide which camera captures the frontal view of the face the best. This camera is labeled the center camera. Third, we select a pair of cameras on both sides of the center camera as a stereo camera and detect the spatial positions of the face and hand. Finally, the target that the subject is pointing to is found on the straight line that connects the face position with the hand position. Experiments show that out system can achieve a mean error of 1.94° with a variance of 4.37 throughout the pointing direction.

1 Introduction

Information expressed by faces and gestures play very important role in human-computer interaction[1, 2]. We are investigating to establish the "Percept-room," which observes people, interprets human behavior, and makes appropriate responses. Since human gestures are key to interpret what a person is doing, many techniques in this research field have been proposed. Among natural human gestures, the pointing gesture is an effective means for a person to notify other people of what that person is interested in. A number of systems have been proposed in the past for human-computer interaction based on pointing gestures[3, 4, 5]. However, they have some restrictions, such as a small range of direction and initialization per user. It is important to detect omni-directional pointing gestures for intelligent environment applications.

In this paper, we propose a multi-camera system that can detect omni-directional pointing gestures and estimate the direction of pointing. In general, a pointing target exists in the direction that a person is facing when they directly point at something (Figure 1). Therefore, we regard the direction of pointing as the direction that is represented by the straight line that connects the face position with the hand position.

2 System Configuration

Figure 2 shows our current system configuration. This system consists of a studio and image input equipment. The studio is a 5 meter × 5 meter square space enclosed by simple backgrounds. We use eight color video cameras placed at an interval of 45° in a horizontal plane with their optical axes crossing at the center of the studio. Each camera is genlocked into a sync generator and the time code is superimposed on their outputs. The image data is input as 640 pixels × 480 pixels size, full color and 30

Figure 1: Example image of a pointing gesture

*Address: 4-1-7 Kagano, Ogaki City, Gifu 503-8569 Japan.
E-mail: {watanabe, hongo, yasu}@softopia.pref.gifu.jp
†Address: 1-1 Yanagido, Gifu City, Gifu 501-1193 Japan.
E-mail: yamamoto@info.gifu-u.ac.jp

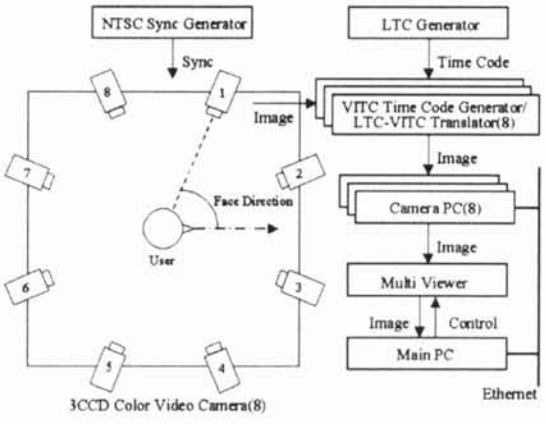


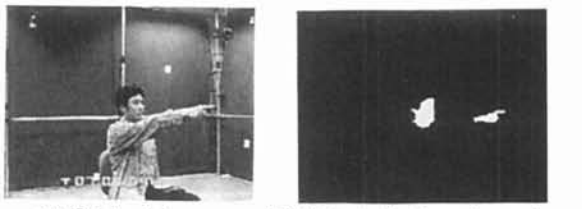
Figure 2: System configuration

frames/sec. and then recorded onto the hard disk in each PC using JPEG compression. The multi viewer combines the individual images into a single image. The main PC can control the multi viewer and choose the input image.

First, the eight cameras detect the face and hand regions using color information. The skin color region in the middle of the image is the candidate for the face region. Then, the face direction is estimated with the discrete face direction feature classes. Second, we estimate the precise face direction with the integrated information from multiple cameras and decide which camera captures the frontal view of the face the best. Third, we select a pair of cameras on both sides of the center camera as a stereo camera and detect the spatial positions of face and hand. Finally, the pointing target is found on the straight line that connects the face position with the hand position.

3 Skin Color Detection

We detect the face and hand regions by skin colors[6]. To extract the skin color area from the input image, we use the LUV color space. First, a two-dimensional UV values histogram is made from the input image. From the two-dimensional histogram, we determine the standard skin color that denotes the maximum number of pixels within the range of skin colors. Second, each pixel's UV value of the input image is converted to the color distance from the standard skin color. Third, we make a histogram of the color distance from the above results and detect the skin color regions by discriminant analysis. Figure 3(a) is an original image, and Figure 3(b) shows the results of the skin color regions. Finally, the skin color region in the middle of the image is the candidate for the face region. The skin regions around the face region are the candidates for the hand region.



(a) Original image (b) Detected skin color regions

Figure 3: Skin color detection

4 Face Direction Estimation

For detecting the omni-directional pointing gestures, we estimate the face directions because the pointing target exists in the face direction. In order to estimate the face direction, we adopt the four directional features[7] that can achieve high recognition rates for the face recognition[8] and the linear discriminant analysis[9].

The four directional features are extracted as follows. First, we make four edge images from the detected face region by applying Prewitt's operator in four directions(vertical, horizontal and two diagonal lines). Second, each edge image is normalized to an 8×8 resolution. An example of the four directional features is shown in Figure 4. Finally, 256 dimensional feature vector is made from these four planes of images. Converting four edge images to low resolution can keep the edge direction information better than directly converting the original image. Consequently, this method is robust against deformation and noise and can be processed quickly.

In discriminating an input vector according to the distance from the mean vector of a discriminant class, it is suitable to compose the classes so that the data in the same class are as near as possible and those in the different classes as far as possible. Under the criteria that the in-class variance is small and the inter-class distance is long, the linear discriminat analysis provides the coefficient matrix $\boldsymbol{A}$ of an optimal linear projection for the training data whose class is known. The feature classes are composed of the linear discriminant analysis of the four directional features that are extracted from some people's faces looking in specific directions.

We have already shown that it is possible to estimate the face direction of an unspecified person by

Figure 4: Four directional features

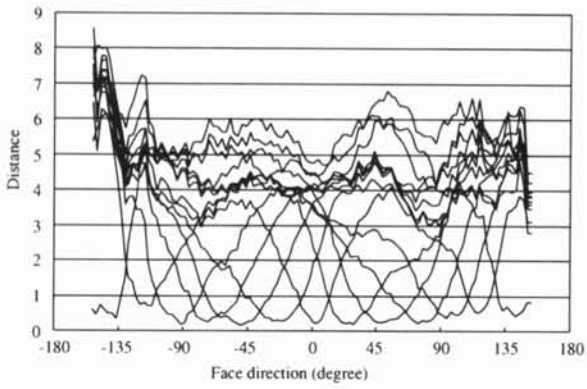


Figure 5: Square distance to face direction class

our proposed method[10]. First, the feature vector $\boldsymbol{x}$ extracted from the face image is transformed to $\boldsymbol{y}$ through the coefficient matrix $\boldsymbol{A}$ that is obtained by the linear discriminant analysis. Second, the distance D_j, between $\boldsymbol{y}$ and the mean vector $\overline{\boldsymbol{y}}_j$ of the face direction class C_j is calculated, where $D_j = |\boldsymbol{y} - \overline{\boldsymbol{y}}_j|^2$. Finally, the direction k that gives the minimum distance is estimated. Figure 5 shows the square distance D_j between the feature vector of a subject's rotating image to C_j.

5 Coordinated Estimation with Multiple Cameras

In the real world, some occlusions between the face and the camera might cause faults in estimation. Therefore, in case of occuring occlusion, in order to get a appropriate estimation for any face direction, we have already proposed the multiple cameras coordinated estimation method[10]. For the purpose of coordinating the information from multiple cameras, the direction component vectors from each camera are integrated into the system coordinates. Then, by equation (1), the face direction j when the evaluation value F gets the largest is estimated.

$$F(j) = \sum_{m=1}^{8} \frac{1}{{D_j}^{(m)} + k} \qquad (1)$$

Where, ${D_j}^{(m)}$ is the distance obtained by the camera m and k is a positive constant for controling amplitude and preventing zero division.

6 Pointing Direction Estimation

To estimate the direction of pointing, we detect the spatial positions of face and hand. The center camera that captures the frontal view of the face is decided from the estimated face direction. Then a pair of cameras on both sides of the center camera

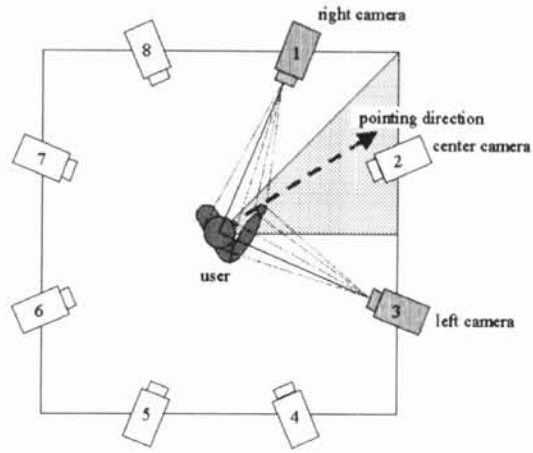


Figure 6: Example of selecting camera

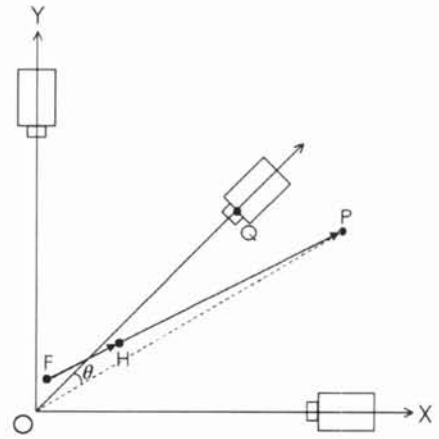


Figure 7: Coordinates for Pointing Direction Representation

are selected as a stereo camera. The selected cameras are shown in Figure 6. Since the two optical axes cross at right angles, the epipolar lines become horizontal and the hand regions are not on the face region. Therefore, we easily estimate the spatial positions of face and hand.

First, the main PC controls the multi viewer and chooses the input image. In this work, two images data that are captured by a pair of cameras on both sides of the center camera compose a image data and are input 320 × 240 pixels size respectively. Second, we detect the face and hand regions by skin colors. the skin color region in the middle of the image is the candidate for the face region. The skin regions around the face region are the candidates for the hand region. Finally, we detect the spatial positions of face and hand. In this work, we used the positions of centers of gravity for the face and hand. Figure 7 shows the coordinates for pointing direction representaion. Where, F is the face position, H is the hand position, P is the target position and the op-

Table 1: Results of experiment

(a) The results of horizontal direction estimation

	−22.5	−11.25	0	11.25
22.5	−18.44	−9.18	2.62	14.16
11.25	−19.40	−10.01	2.85	12.92
0	−20.73	−9.46	1.49	13.05
−11.25	−20.52	−10.13	0.60	12.74
−22.5	−21.13	−9.03	0.62	12.17

The mean error : 2.14 The variance : 2.90

(b) The results of vertical direction estimation

	−22.5	−11.25	0	11.25
22.5	22.00	21.56	21.72	20.72
11.25	10.81	11.33	12.10	11.40
0	−1.16	0.01	0.17	1.29
−11.25	−12.26	−10.66	−9.21	−8.11
−22.5	−23.73	−21.58	−19.73	−18.28

The mean error : 1.74 The variance : 4.79

tical axis of the camera Q passes through the origin O. Since P exists on the line that extends FH, the pointing direction θ is the angle between the camera Q's optical axis and the line OP.

7 Experiment

In this experimental data, subjects look at specific targets and point to them. The result shows the estimated angle to 20 targets that were arranged by four kinds in the horizontal directions ($-22.5°$, $-11.25°$, $0°$ and $11.25°$) and by five kinds in the vertical directions ($-22.5°$, $-11.25°$, $0°$, $11.25°$ and $22.5°$). When the target is the camera Q, the angle is the horizontal $0°$ and vertical $0°$. Table 1 shows experimental results for 14 subjects. In table 1(a), the values in the first row of the table indicate the correct angles and those in rows 2–6 indicate the estimations in the horizontal direction. The mean error of the horizontal direction was $2.14°$ and its variance was 2.90. In table 1(b), the values in the first column of the table indicate the correct angles and those in columns 2–5 indicate the estimations in the vertical direction. The mean error of the vertical direction was $1.74°$ and its variance was 4.79. The whole mean error was $1.94°$ and the whole variance was 4.37. This result demonstrated that our proposed method could estimate the direction of pointing.

8 Conclusion

We developed a system that can extract the omni-directional pointing gestures and estimate the direction of pointing. To detect the omni-directional pointing gestures, the eight cameras estimate the face directions because the pointing target exists in the face direction. Then, to estimate the direction of pointing, a stereo camera detects the spatial positions of face and hand because a pointing target exists in the direction that is represented by the straight line that connects the face position with the hand position. In this method, therefore, the initialization per user is unnecessary. As the experimental results, the mean error was $1.94°$ and its variance was 4.37. Our next tasks will be to improve the precision of the pointing direction estimation.

References

[1] M. Kaneko and O. Hasegawa, "Processing of face images and its applications," IEICE Trans. Information and Systems, vol.E82-D, no.3, pp.589-600, 1999.

[2] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.22, no.1, pp.107-119, 2000.

[3] M. Fukumoto, et al., "Realtime detection of pointing actions for a glove-free interface," In Proc. IAPR Workshop on Machine Vision Applications, pp.473-476, 1992.

[4] R. Cipolla, et al., "Uncalibrated stereo vision with pointing for a man-machine interface," In Proc. IAPR Workshop on Machine Vision Applications, pp.163-166, 1994.

[5] N. Jojic, et al., "Detection and estimation of pointing gestures in dense disparity maps," In Proc. of 4th International Conference on Automatic Face and Gesture Recognition, pp.468-475, 2000.

[6] H. Hongo, et al., "Focus of attention for face and hand gesture recognition using multiple cameras," In Proc. of 4th International Conference on Automatic Face and Gesture Recognition, pp.156-161, 2000.

[7] K. Yamamoto, "Present state of recognition method on consideration of neighbor points and its ability in common database," IEICE transactions, vol.E79-D, no.5, pp.417-422, 1996.

[8] S. Kuriyama, K. Yamamoto, et al., "Face recognition by using hyper feature fields," Technical Report of IEICE Pattern Recognition and Media Understanding, vol.99, no.448, pp.105-110, 1999.

[9] T. Kurita, "A study on applications of statistical methods to flexible information processing," Researches of the Electrotechnical Laboratory, no.957, 1993.

[10] M. Yasumoto, H. Hongo, et al., "A consideration of face direction estimation and face recognition," In Proc. of Meeting on Image Recognition and Understanding, vol.1, pp.469-747, 2000.