Face-Direction Estimating System Using Stereo Vision

Takuya MINAGAWA Hideo SAITO Shinji OZAWA

Department of Electrical Engineering, Keio University

3–14–1 Hiyoshi Kouhoku-ku Yokohama 223, JAPAN

TEL: +81-45-563-1141 (ext.3309) FAX: +81-45-563-2773

{takuya, saito, ozawa}@ozawa.elec.keio.ac.jp

Abstract—**We developed the system which can estimate the face-direction of a person with two cameras using stereo vision. In this system, the eyes and the mouth are extracted in each image for measurement of their 3D position by matching the each feature in the one image with in the other. Then, face direction is obtained from the normal of the plane which is determined by the 3D coordinates of the features. This system can track face-direction in semi-realtime.**

I. Introduction

Estimation of face direction is one of the most important technology for useful man-machine interface. For example, Ballard and Stockman proposed the system that facial aspect is used to control the cursor[1]. For the purpose of tele-conference, facial movement estimation has been studied to send codes at high speed[2]. Human gaze estimation also needs the information of face direction[3].

In most cases, the estimation of face direction requires a lot of time, or needs the information about the position of features. In this paper, we propose semi-realtime system which can estimate human face direction. Using stereo vision, this system does not need to fix the distance between the features.

II. Outline of the System

The outline of the system is shown in Fig.1. The position and orientation of a face can be computed from the 3D coordinates of three feature points: the eyes, and the mouth. The 3D coordinates of the each point are estimated by stereopsis images of two cameras placed in front of a user. These two cameras are connected to Hitachi IP-X image processing board which can compute filtering, correlation, histogram, labeling, etc in about 11 ms each, and which is loaded on PC with 486DX2(66MHz) and ODP.

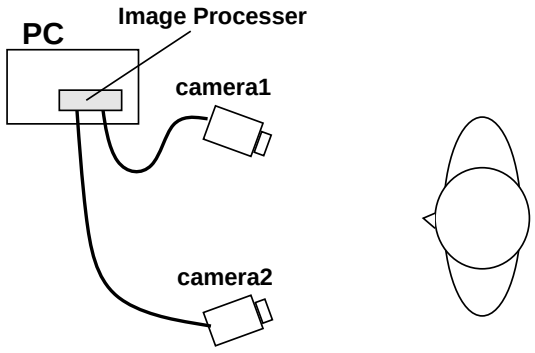


Fig.1. Outline of the system

Because it is easier to estimate the face direction, these two cameras are not parallel, which make their images difference about the distance between the features.

III. Process of the System

As mentioned above, the 3D coordinates of the each feature can be computed by the stereometric ranging. Therefore, the position of the feature should be found in the images of camera 1 and camera 2. At the first, a user must expose its full face to the camera 1. Then, the user is allowed to move its face as he or she likes. The flow of the system process is shown in Fig.2.

First, the images are taken with the both cameras. Then, in the image of camera 1, the eyes are extracted using template matching. With reference to this eyes' position, the mouth position is extracted. These three features in the image 1 are used to locate the matching parts in the image of camera 2. Finally, the system calculates the 3D position of the each feature and the face direction.

0-7803-3932-0

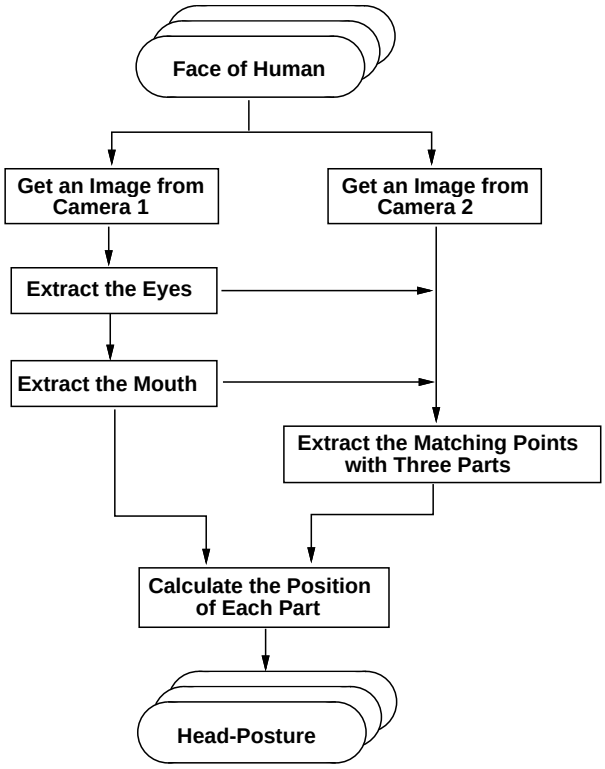
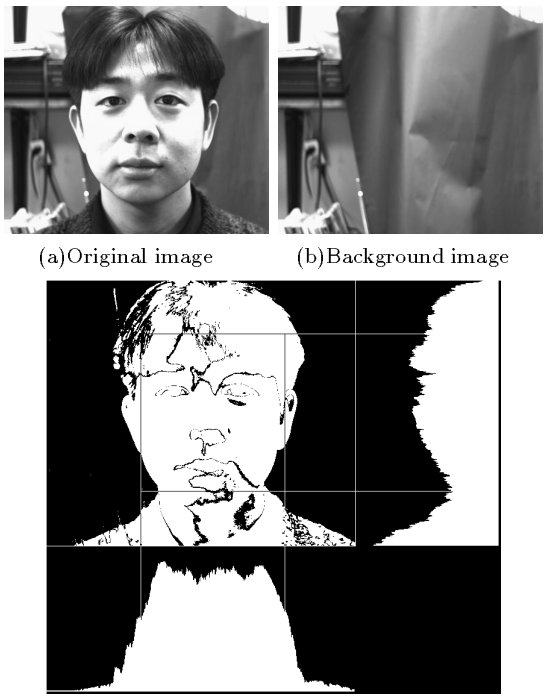


Fig.2. Process of the system



(a)Original image (b)Background image

(c)Subtracted image

Fig.3. Extraction of the head

A. Preprocessing

First of all, the initial positions of eyes are detected from the image of camera 1 as preprocessing. Mukai, Mitani, and Togaawa used the edge information to locate the eyes[4]. However, this technique has probabillity to extract wrong feature like eyebrow or hair. Then, we extract the head position before the eye extraction.

1) The Estimation of the Head's Position

The head's position is extracted from the feature of it's shape. The shape of the head can be distinguished for subtracting background image from human image and binarizing this. This subracted binary image is projected to X and Y axis. As shown in Fig.3, the projected points on X axis may have large value in the range where the head are. Thus, the range over threshold level is extracted as the width of the head.

About the projected value on Y axis, the top of the head is searched from the top of the image and extracted the first point that is over threshold level. Since the points on Y axis may have minimum value around the neck, the system searchs the smallest point on Y axis. To avoid finding the smallest point around the top of the head, the search starts from the point that is half the width lower than the top of the head.

2) The Estimation of the Eyes' Position

In the image of human face, there is uneven and rapid change of gray level around the eye and the eyebrow. For that reason, the eyes and the eyebrow may have a lot of edges. Using the distinctive feature of the edge around the eyes, we extract the eyes' height and the horizontal position separately.

The eye extraction is illustlated in Fig.4. The process below is executed in the extent of the head. First, the face image is smoothed to remove noise. Then, the edges are enhanced with Robinson operator, and this image is binarized to extract the edges. In the Y axis that the edge image is projected to, the system extracts the ranges, which are over the threshold level, as candidates of the eye. The largest value in the each candidate are compared each other, and the ranges are selected that have the largest and the second largest velue as the eyes and the eyebrows. Then, the lower positioned range is selected as the eyes' height.

In this range of the eyes height, the binarized edge image is projected to X axis. As well as the Y axis, the ranges over threshold level are extracted as the candidate of the eye. In these candidates, two of them are chosen, which have the most appropriate width, judging from the ratio of the eye's length to width.

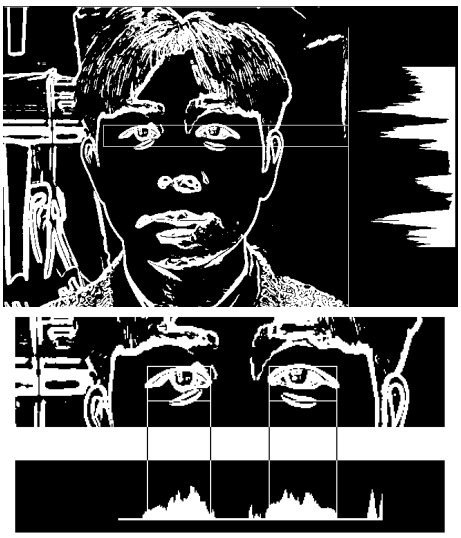


Fig.4. Extraction of the initial eye position

B. The Feature Extraction

After the preprocessing, the system gets the images with both two cameras. In the image of camera 1, the three features – the eyes and the mouth – are detected. As well as the preprocessing, the edges of the image 1 are extracted with Robinson operator and binarization. This edge image is used in following process.

1) The Eyes Extraction

First, a template is determined in the previously made templates, to be the most similar size to the eyes that obtained with the preprocessing. The template is shown in Fig.5. Its contour is formed with two arcs. Inside the contour, it is not used to compute the correlation, because the pupil can move in the eye. Outside the contour, there is margin, because an eye may be isolated from any other features.

The eyes are extracted using the information of the eye position on the previous frame. Around the eyes on the previous frame, the window is set; the eyes are searched in this window with normalized cross correlation.

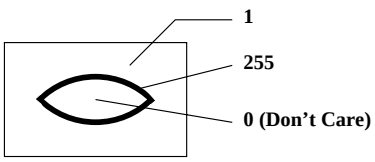


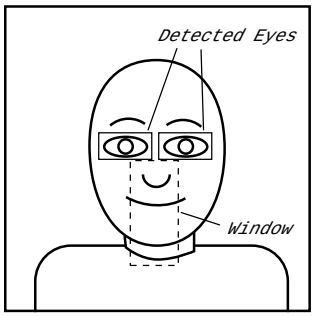
Fig.5. The template of eye

2) The Mouth Extraction

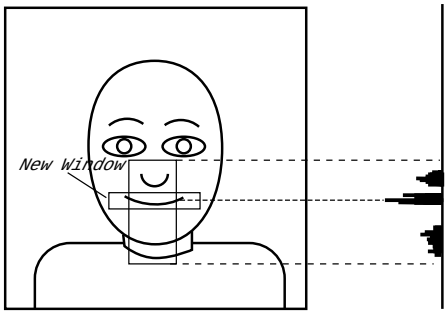
The mouth is detected to be the area that has the most concentrated edges under the eyes. Thus, the system executes the following setps: (1)setting the window with reference to the eyes (Fig.6 (a)), (2)projecting the edge image in this window to Y axis, (3)finding the position that has the largest value on the projected histogram as the height of the mouth, (4)setting the new window around the height of the mouth(Fig.6 (b)), (5)projecting the image in this new window to X axis, (6)extracting the range that is over threshold level as the mouth.

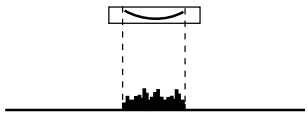
C. The extraction of the matching feature points

After the three features are extracted in the image 1, the similar parts are searched in the image of camera 2. First, the binarized edge image is made from the image 2 with Robinson operator and binarization. Secondly, the extracted feature in the image 1 is entered as the template. Thirdly, the search window is set in the image 2. And finally, template matching is executed in the window to find the corresponded feature with the template.



(a) Determination of the window





(b) Extraction of the mouth with projection

Fig.6. The extraction of the mouth

The features in image 2 are extracted in order of the right eye, the left eye, and the mouth. The search window is determined with the epipolar line and the features extracted before: the search window of the left eye, for example, is the rectangle placed at the right hand of the right eye (Fig.7).

D. Computing the face direction

The 3D coordinate of the each feature point can be computed with the 2D coordinate of the feature in image 1 and image 2. As shown in Fig.8, f is supposed to be the focal length of the each camera, and d is supposed to be the distance between two cameras. The point used in calculation is the center of the extracted feature rectangle. The $p_l(x_l, y_l)$ is asuumed to be the feature point on the image plain L ,and $p_r(x_r, y_r)$ on the image plain R is asuumed to be the matching point with p_l. Then, the 3D coordinate of this point $P(X_P, Y_P, Z_P)$ is determined by following equations:

$$X_P = \frac{(x_l y_r + x_r y_l)\sin(\alpha) + (y_r - y_l) f \cos(\alpha)}{(x_l y_r - x_r y_l)\sin(\alpha) + (y_r + y_l) f \cos(\alpha)} \cdot \frac{d}{2}, \quad (1)$$

$$Y_P = \frac{y_l y_r d}{(x_l y_r - x_r y_l)\sin(\alpha) + (y_r + y_l) f \cos(\alpha)}, \quad (2)$$

$$Z_P = \frac{(f \sin(\alpha) + x_r \cos(\alpha)) \cdot y_l d}{(x_l y_r - x_r y_l)\sin(\alpha) + (y_r + y_l) f \cos(\alpha)} \quad (3)$$

$$= \frac{(f \sin(\alpha) - x_l \cos(\alpha)) \cdot y_r d}{(x_l y_r - x_r y_l)\sin(\alpha) + (y_r + y_l) f \cos(\alpha)}, \quad (4)$$

where α is the angle of the camera. If the Eq.(3) and Eq.(4) were not equal, there would be no intersection between the two lines that runs from the camera through the image plain.

If the 3D positions of the three features are obtained, the face direction can be calculated as normal vector of the plain of a human face. The position vectors of three features are assumed to be:

$$A = (x_1, y_1, z_1)^T, \quad (5)$$

$$B = (x_2, y_2, z_2)^T, \quad (6)$$

$$C = (x_3, y_3, z_3)^T. \quad (7)$$

Then normal vector $\boldsymbol{n} = [X_d, Y_d, Z_d]^T$ can be computed by the following equations.

$$X_d = (y_2 - y_1)(z_3 - z_1) - (y_3 - y_1)(z_2 - z_1) \quad (8)$$

$$Y_d = (z_2 - z_1)(x_3 - x_1) - (z_3 - z_1)(x_2 - x_1) \quad (9)$$

$$Z_d = (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) \quad (10)$$

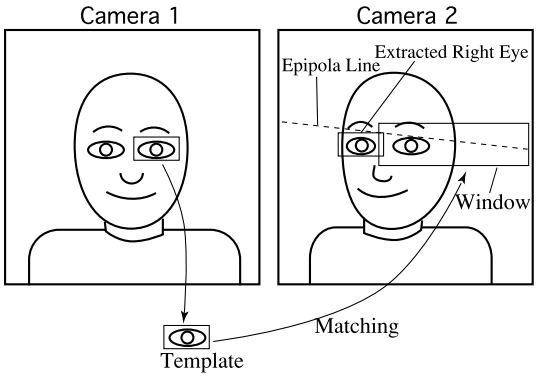
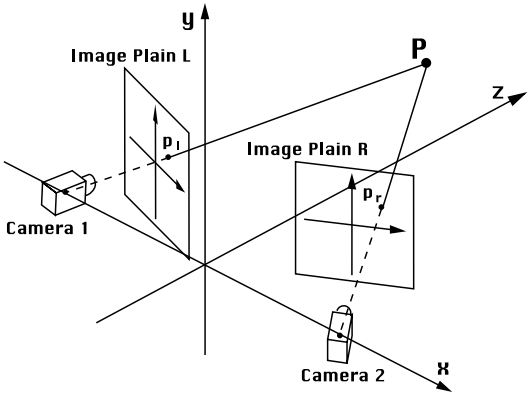
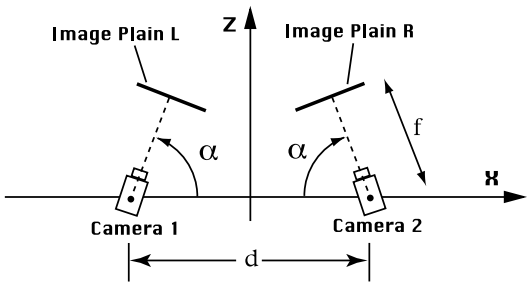


Fig.7. Extraction of the left eye in image 2



(a)



(b)

Fig.8. Stereoscopic imaging

IV. Experimental Result

As mentioned before, the cameras are inclined in order to make the estimation more accurate. However, if the inclination of the cameras is too big, the difference between two images would also be too big to get the matching feature among the images.

Considering this trade-off, we decided the each parameter in Fig.8 as follows:

Angle of the cameras, α: 80°
Distance between the cameras, d: 23cm
Focal length of the camera, f: 1.25cm .
The time spent for the processing is shown in Table I.

For the evaluation of this system, the user was asked to move his head slowly; the motion of this head was recorded on the hard disk with the two cameras. We investigated the result of the each process using these images which are consisted of 512 × 440 pixels, and have 256 gray level.

A. The Evaluation Using the Model

We used the human face model which can be sythesized its direction. In order to evaluate the computed direction visually, the model was changed to the direction estimated by the original image, and compared with the original. Fig.9 shows a example of the original image and the synthesized image.

TABLE I
PROCESSING TIME

Extraction of the initial eye position	Once Estimation	
	Camera switching	Process
0.24 sec	0.3×2 sec	1.28 sec

(a) The original image

(b) The synthesized image

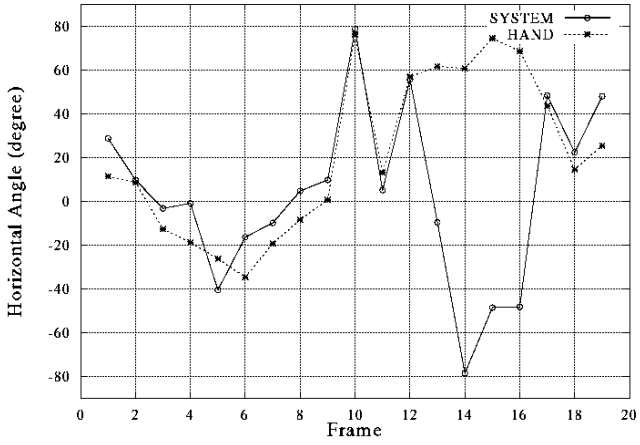
Fig.9. Comparison with the model

B. The Evaluation Using the Feature Points Extracted by Hand

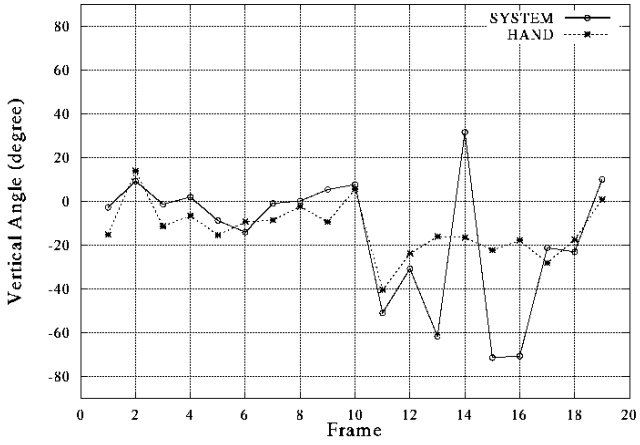
On the images of camera 1 and camera 2, we extracted the feature points and calculated the direction of the face by hand. The direction estimated by the system was compared with the estimation by the hand, which is shown in Fig.10.

V. Discussion and Conclusion

In this paper, we have presented the system that estimate the human face direction using stereo vision. This system extracts the eyes and the mouth in the image of two cameras; the face direction is computed with the position of each feature in this two images.



(a) The horizontal angle of the head



(b) The vertical angle of the head

Fig.10. Angle of the head on each frame

As the result of the experiment, it is turned out that this system has some advantage:

1. The initial eye postion can be extracted robustly, because the information of the head position is used to limit the search area.

2. The eye can be tracked robustly with the template which has the margin, because an eye is isolated from any other features.

3. The face direction can be estimated without any information about the distance among the features, because stereo vision can determine the 3D position of the feature point.

VI. References

References

[1] Philippe Ballard,George C Stockman:"Controlling a Computer via Facial Aspect"!$*IEEE Transactions on Systems,Man,and Cybernetics*!$Vol.25!$No.4!$ pp.669-677 (1995)!%

[2] Haibo Li, Robert Forchheimer:"Two-View Facial Movement Estimation", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.4, No.3 ,pp.276-287(1994)

[3] Andrew Gee, Roberto Cipolla: "Determining the Gaze of Faces in Images", *Image and Vision Computing*, Vol.12, No.10, pp.639-647(1994)

[4] Toshio Mukai, Junji Mitani, Fumio Togawa: "The Method of getting Gaze Direction using Image Processing", *Proceedings of the 2nd symposium on Sensing via Image Information(in Japanese)*, pp.135-138 (1995)