Person Identification by Multi-Directional Face Recognition while Walking in a Room

Suriyon Tansuriyavong † Jun-ya Ebisawa † Masayoshi Endo † and Shin-ichi Hanaki †
† Department of Electrical Engineering Nagaoka University of Technology
1603-1 Kamitomioka - machi, Nagaoka, 940-2188 JAPAN
E-mail: suriyon@vos.nagaokaut.ac.jp

***Abstract*— Person tracking by image processing has been used for human monitoring in a room. Human image is extracted from a picture by using image difference between video-picture frames. Also, face recognition technique is employed at the room-entrance to identify and give the name to the person. However, a person may not be identified for several reasons such as "rejection" in the face-recognition at the room-entrance, stepping out of view temporarily, or failure of resuming original name after image-occlusion caused by multiple person-image overlapping. To improve name-confirming opportunities, this paper proposes a multi-directional face-image recognition technique while persons are walking around in a room. The proposed system employs additional active camera besides a conventional fixed camera. When the fixed camera detects a person who moves toward the active camera, it zooms into the person's face to obtain the zoomed-up face image, which is recognized by using the template matching technique. For captured face images, the recognition rate was 85.5%.**

***Index Terms*—human monitoring, awareness, person tracking, face recognition, face image acquisition.**

I. INTRODUCTION

In order to supply awareness information, video systems that always send video image of a room have been investigated in many studies[1][2][3]. Although sending incessant video image is useful for supporting awareness at the receiving side, it causes privacy infringement issues. To protect subject privacy while supporting useful awareness information, a system was proposed that extracts human image by making use of video frame difference and that identifies person's name by face recognition using a camera located at the room entrance, tracks the person's image by another video camera in the room, and conceals person's image with silhouette [4]. It uses a camera at the room entrance, because most face recognition methods so far have required the front face image with little margin. In most works, front face image is obtained from a stationary person in an experimental environment. An attempt was made to obtain the front-face image from a walking person at the room-entrance by providing a camera that is set at the corner of a specially arranged L-shape lane [4].

Target image tracking with person's ID (identification) propagation technique was employed. However, name as ID is unavailable by several reasons such that the face-recognition system can give no specific name even to the registered-person at the room-entrance, the target person temporarily goes out of the scene and later comes back, or resuming original name fails after image-occlusion caused by the plural person-image overlapping.

Meantime, several studies have been made to allow not only the front face but also the multi-directional face recognition. They have been limited, however, to the experimental environments where the background and lighting are controlled.

The present study aims to increase recognition opportunities by using multi-directional face image instead of traditional face-recognition technique that utilizes only the front face, which is made available both by tracking and capturing face image while persons are walking in practical room environments. It also obtains face-recognition results for each video frame by employing simple-but-fast pattern matching method in order to choose those valid video-frames for face-recognition. In the proposed system, acquisition of the multi-directional face image is realized even in a complicated background, by using skin color information, multi-directional averaged face templates (MDAF templates), and the position of face organs.

The system consists of a video camera so that it is set to monitor the wide area of a room, an active camera that has pan, tilt and zoom capabilities, and a personal computer (PC) for image processing and camera controlling. The major functions of the system include the person tracking, the person's head part image shooting, the face image acquisition, and the multi-directional face recognition.

In order to build multi-directional averaged face templates (MDAF templates) and face dictionary for recognition, face image data whose direction is precisely known are needed. For the purpose, a multi-directional face image acquisition system was built.

II. FACE DATA ACQUISITION

A. Multi-directional face image acquisition system

The configuration of the multi-directional face image acquisition system is shown in Fig.1.

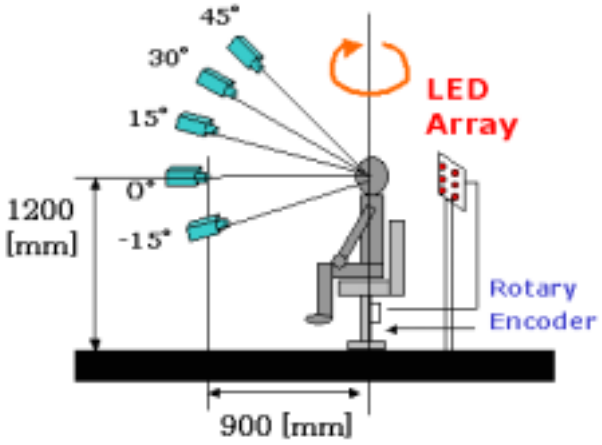


Fig.1 The configuration of multi-directional face image acquisition system.

Five fixed cameras are arranged on a semi-circle in front of a rotatable chair. The face image is continuously taken as the chair rotates, on which a subject sits.

Because cameras are set near the ceiling in the actual environments, most cameras of the experimental system were located at a higher position than subject face.

An optical-rotary-encoder is installed on the chair, which is used to know the rotation angle of the face in the picture taken by each camera. An LED array to indicate rotation angle in Gray code bit patterns is arranged, whose image is recorded with face images at the background.

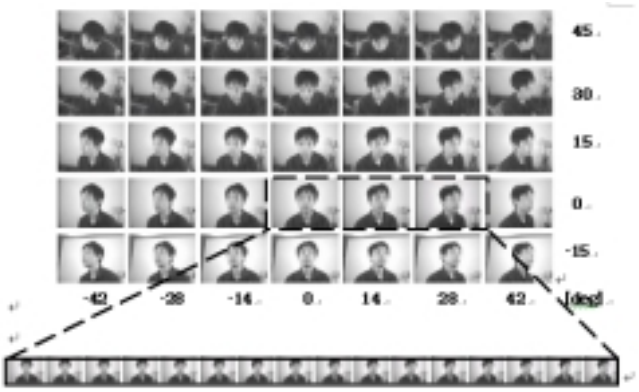


Fig. 2 The sample of multi-directional face data.

A set of 5 video cameras and 2 PCs made it possible to acquire multi-directional face data in a short time. These face data are stored in AVI (Audio Video Interleaved) format without compression.

The multi-directional face data from 11 subjects were acquired by 15-degree interval in the vertical and 2-degree interval in the horizontal directions. Although each camera supplies face image at 2-degree interval as shown at the bottom of Fig.2, only face images at 0°, ±14°, ±28°, and ±42° are selected as representative. A sample of multi-directional face data is shown in Fig. 2.

B. Template for detecting face image

Face images that belong to the identical direction group are averaged to obtain face-template of that direction, which is later used to detect faces in a video image shown in Fig.3 (a) by the technique proposed by Kosugi[5]. The face region in an image is decided based on the position of both eyes and mouth. For every subject, multi-directional face samples are obtained as shown in Fig.3 (b).

Then, face images are averaged for all subjects, at each specific direction from which mosaic-images of the averaged face is made as shown in Fig.3 (c). Here, “mosaic processing,” means resolution reduction through pixelization. Finally, the face-background and the hair regions are removed by manual operation as don't-care-region. The MDAF templates are made form 11 persons in the size of 12×12 as shown in Fig.3 (d).

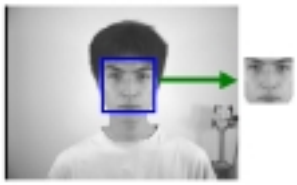
III. MOVING-PERSON IDENTIFICATION BY FACE RECOGNITION

A. System configuration

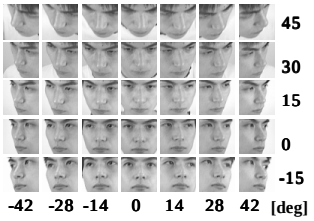
The configuration of moving-person identification system is shown in Fig. 4. The system consists of a fixed camera for person tracking, an active camera for face image acquisition and a personal computer for camera control, image processing and face recognition.

In order to realize faced recognition with high recognition rate, a fine face image is required. Small face image in a room scene is not enough for face recognition and zoomed face image with high resolution is preferable. Therefore, an active camera to shoot face image zoomed-up is necessary.

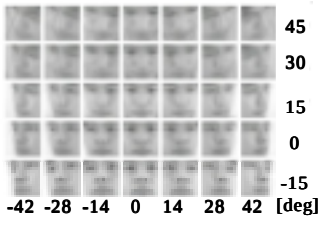
The basic idea of this system is based on the behavioral tendency, i.e. "A person tends to turn face toward the direction which he/she bounds for." The image taken by the fixed camera in Fig. 4 is used to estimate the position and the moving direction of the person. It is assumed that person’s feet-point can be always observed from the fixed camera. The person who moves toward the active camera is selected as a candidate to be recognized. The active camera is targeted to the subject head, and zoomed so that the face image becomes suitable size.



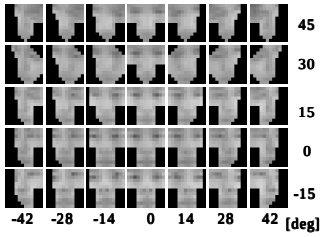
(a) The decision of the face region.



(b) Multiple-directional face data.



(c) Mosaic image of multi-directional averaged face images.



(d) Multiple-directional averaged face templates (MDAF templates).

Fig.3 The preparation of the MDAF templates.

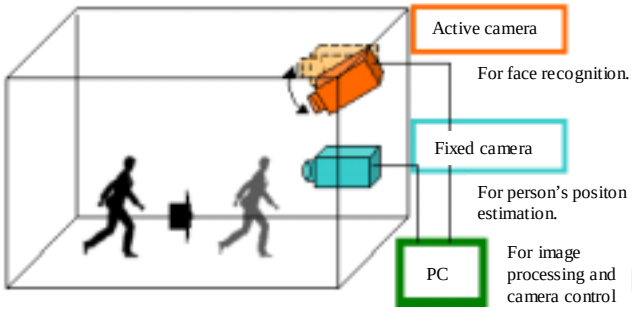


Fig. 4 The configuration of moving person identification system

This system was implemented with personal computer having Pentium Ⅳ CPU whose clock frequency is 2.0 GHz, with 512 MB memory and operating under the Windows 2000.

B. *The processing flowchart*

The processing flowchart is shown in Fig. 5. The head image is acquired while tracking the person moving toward the active camera. From the head image, face image is determined. The feature vector is extracted from the face image, which is compared with standard-face vector of each registered person in the face dictionary to obtain recognition result.

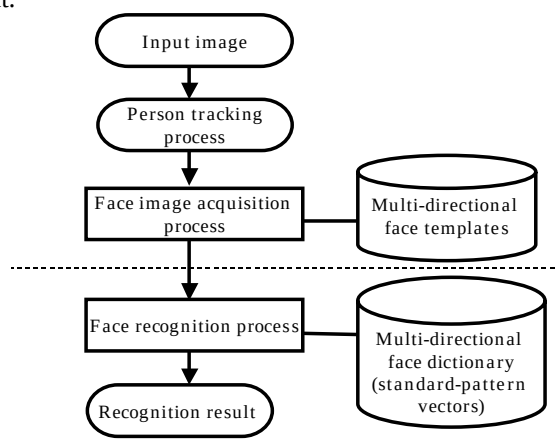


Fig. 5 The processing flowchart.

C. *Face image acquisition process*

In a room scene, candidate face region is extracted by using skin color information. The extracted skin colored region is quantized and used for determining face-searching area. Because, the exact face size is not known, the skin color region is arranged in three different sizes as shown in Fig.6 (d). MDAF templates are used to scan over the skin color region in order to surely find out possible face image.

Both eyes are detected by using the gradient of gray image. Mouth is also detected by red color region in the face-searching area.

The matching value for each template must be higher than a predetermined threshold, and the region whose matching score is highest is considered as the face region. If all templates fail to give an adequate matching score, then it is determined that no face image exists in the region.

D. *Face recognition process*

Feature vector is determined whose 144 components have pixel value of mosaicked face image (12×12 blocks). By using mosaicked image, the recognition algorithm is less affected by facial hair or with or without glasses. The distance is calculated between feature vector of the input image and standard-pattern vector registered in the face dictionary. For a feature vector, the name of the nearest standard pattern is assigned as the recognition result.

Euclidean distance between the pattern $\mathbf{p}_i$ and $\mathbf{p}_j$ both expressed as vectors is defined as equation (1), which is used as the measure between the patterns.

$$D_E(\mathbf{p}_i, \mathbf{p}_j) = \sqrt{(\mathbf{p}_i - \mathbf{p}_j)^t (\mathbf{p}_i - \mathbf{p}_j)} \qquad (1)$$

If the minimum distance to the nearest standard-pattern exceeds criteria, the input pattern is rejected.

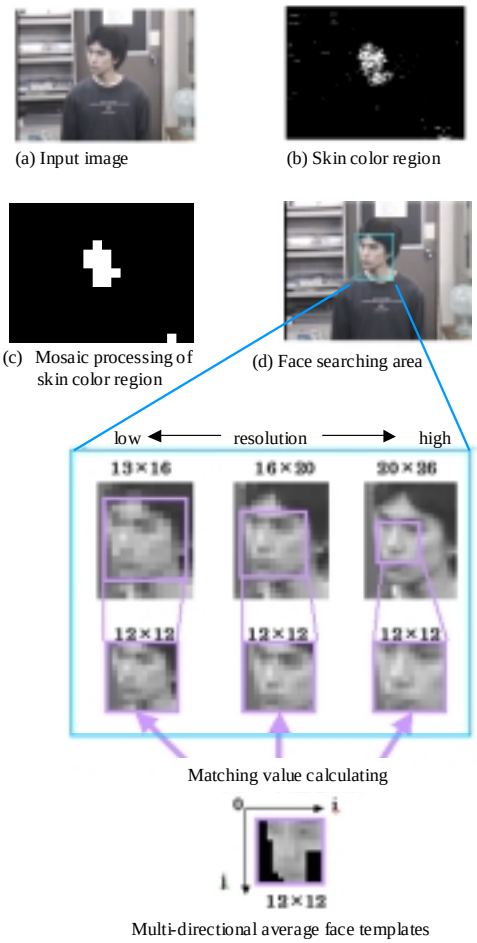


Fig. 6 The decision of face region in the input image

IV. EXPERIMENT

A. Experimental environments

An evaluation experiment was carried out using implemented system to acquire and recognize face image of person moving in the room.

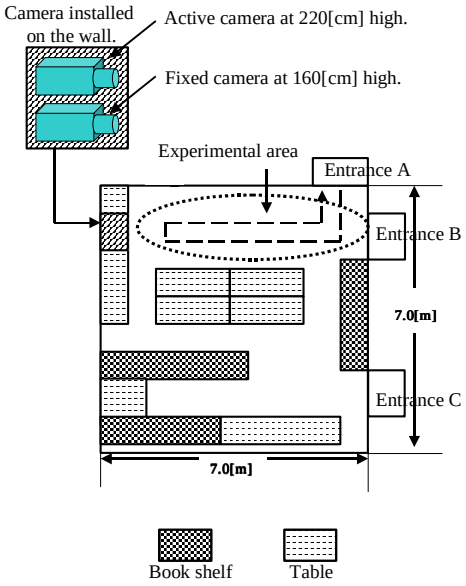


Fig. 7 Experimental environments.

Experimental environments are shown in Fig. 7. Both fixed camera and active camera were installed on the wall. The fixed camera was set at 160 [cm] high, and the active camera was set at 220 [cm] high.

Experimental subjects were asked to walk from the entrance A, follow the route and go out from the entrance A. They were also asked to change their face direction side by side while walking through the experimental route, which helps for the system to acquire face image data in various directions more efficiently during the walk.

B. Evaluation of face image acquisition using MDAF templates

Five MDAF templates including the front face template were used for extracting face images after evaluation of each template coverage.

In Table1, the number of frames successfully acquired face image at each direction and their total frame number, 1296[fr] are shown. The front face template at the direction [0,0] also covers the template at the direction [0,30].

Table 1 The total number of frames that face images were acquired

Vertical direction [degrees] \ Horizontal direction [degrees]	-30	0	30	Total
30	168	356	193	361
0	265		314	935
Total	433	356	507	1296

The number of frames acquired by using the front face template only.

The number of frames acquired by using MDAF templates.

By using MDAF templates a total of 1,296 frames were acquired. It implies that a face recognition opportunity increased by 3.6 times, comparing with the case using front face template only. Samples of face images successfully acquired in various directions are shown in Fig. 8.

Fig. 8 Samples face images in various directions.

C. Face recognition experiment

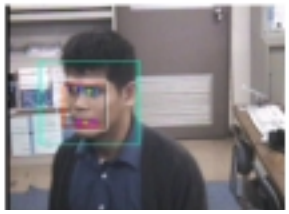
Face images acquired were used for the face recognition evaluation. As the accurate face direction is not known beforehand, all possible face templates were applied and the direction of the template that gave the highest score was assigned as the nominal direction of the face. The standard-pattern vector in the face dictionary of the same direction is used for recognition process.

The average recognition rate for 11 subjects was calculated for each direction and the results are shown in Table2. The average recognition rate for all direction was 85.8%.

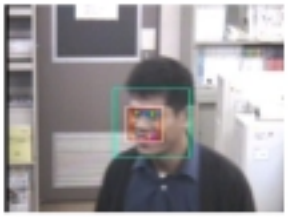
Table 2 The face recognition rate from experiment

Vertical direction [degrees] \ Horizontal direction [degrees]	-30	0	30	Average
30	84.9	90.7	83.2	84.1
0	86.0		84.4	87.0
Average	85.5	90.7	83.8	85.8

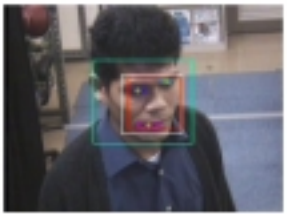
Face recognition examples are shown in Fig. 9. The outer rectangle in Fig.9 shows the searching area of face, and the inner rectangle shows the face area decided by average face template.



(a) A sample of correct recognition



(b) A sample of mis-recognition due to varing of facial expression



(c) A sample of mis-recognition due to varing of lighting

Fig. 9 The sample result of face recognition.

Figure 9 (a) shows correct recognition case, in which both eyes and mouth are correctly detected. Fig. 9 (b) shows recognition failure due to the large discrepancy in facial expression. In the case of Fig. 9 (c), failure was caused by unexpected shading patterns. It can be improved by adding standard-pattern vectors with corresponding lighting.

Processing time was 50 [msec] from the image capturing process to the face recognition process.

V. CONCLUSION

A moving-person identification system is proposed and realized. The system involves a fixed video camera for person tracking, and an active camera for face image acquisition. The active camera is controlled so that face image is obtained as fine as possible in order to achieve face recognition with a high recognition rate.

A method is proposed that the active camera shoots those who moves toward it, and that the head portion of the person is zoomed-up and its image is obtained by the system. Those who moves toward the active camera is detected by tracking the person from the fixed video camera image.

It provides a method of acquiring face image by chance from those moving in the practical room environment.

Multi-directional averaged face templates (MDAF templates) were devised to detect not only the front face but also the multi-directional face for recognition targets.

By using MDAF templates, face recognition opportunity increased by 3.6 times, comparing with the case of using front face template only.

Muti-directional face dictionary was built, which was used to examine feature vectors obtained from acquired face image.

The face recognition rate was 85.8%.

It seems to be useful to recover the person's ID after "rejection" in the face-recognition at the room-entrance, stepping out of view temporarily, or failure of resuming original name after occlusion in incessant video image for supporting awareness information.

Position estimation will be substituted by the stereovision method, which covers even the case where feet-position can not be seen, with some extra computing cost.

ACKNOWLEDGMENT

Authors thank Mr. Shimoda Masaaki who first developed the base system just capturing front faces. They also thank all the students of the laboratory who have cooperated to experiments in this research.

REFERENCES

[1] Robert Fish et al.：" The VideoWindow System in Informal Communication," ACM Conf. Computer-Supported Cooperative Work (CSCW'90), ACM, pp.1-11, (1990).

[2] Dourish, P., and Bly S., "Portholes: Supporting Awareness in a Distributed Work Group," CHI '92, May 3-7, pp.541-547 ,(1992)

[3] Shin-ichi Hanaki and Suriyon Tansuriyavong: "A Study on Room Computer having a Field that observes Users or Group Members (in Japanese)," Technical report of IEICE, MVE2000-1, pp1-6, (2000).

[4] Suriyon Tansuriyavong and Shin-ichi Hanaki："Privacy protection by concealing persons in circumstantial video image, " Workshop on Perceptive User Interfaces (PUI'01), (2001).

[5] Makoto Kosugi and Kouji Yamashita: "Person Identification System Based on a Trapezoid Pyramid Architecture of a Grey-Level Image," ICIAP pp. 501-508, (1997).