Development of vision based meeting support system for hearing impaired

R Shikata[1], T Kuroda[2], Y Tabata[3], Y Manabe[4] and K Chihara[4]

[2]Department of Medical Informatics, Kyoto University Hospital,
Kawaramachi, Shogoinn, Sakyo, Kyoto, JAPAN

[3]Department of Radiation Technology, Kyoto College of Medical Technology,
Sonobe, Nantan, Kyoto, JAPAN

[4]Graduate School of Information Science, Nara Institute of Science and Technology,
Takayama, Ikoma, Nara, JAPAN

[1]*ryo@7501.net,* [2]*Tomohiro.Kuroda@kuhp.kyoto-u.ac.jp* , [3]*yoshi-t@kyoto.medtech.ac.jp,*
[4]*manabe@is.naist.jp*, [4]*chihara@is.naist.jp*

[2]*www.kuhp.kyoto-u.ac.jp,* [3]*www.kyoto.medtech.ac.jp,* [4]*chihara.naist.jp*

ABSTRACT

This paper describes a new meeting support system that helps the hearing impaired to understand the contents of the meeting. The proposed system distinguishes the mainstream of the discussion from other chattering based on the utterances. The situation of the meeting is acquired as a picture using an omni-directional vision sensor, and the system analyzes speaker's relations from the captured image by using face directions for the participants. The system shows the mainstream and the chattering of a meeting by using the analyzed result and speech-recognition.

1. INTRODUCTION

In a meeting, the hearing impaired would have difficulty in catching the meeting's contents without note-taker or sign interpreters. The participants in the meeting have some opinions and discuss with another participants. When a participant speaks the opinion, hearing people find the speaker from direction of voice, but the hearing impaired are difficult. The hearing impaired catches the meaning of a speech by lip-reading, when they discuss with a hearing person. The lip-reading is an effective way while they talks to hearing person face to face. Therefore, the hearing impaired has trouble to finding the speaker at the meeting of a large attendance and has trouble to understanding the contents.

There are two ways to let the hearing impaired know contents of meeting. One is to help by sign interpreters or note-takers. The other is to help by using computer system. However, the former has problems that it takes a long time to train sign interpreters or note-takers, and that the shortage of the interpreters occurs. As the latter, the basic technologies like speech recognition have been developed and have been utilized in some kinds of fields. To make the meeting support system, the speech recognition is an effective technology, but most speech recognition method is utilized against a speaker. Thus, it is necessary to have an idea to use the developed speech recognition in the meeting of a large attendance.

This paper proposed a new support system for the hearing impaired. The proposed system captures the scene of meeting and voice data of participants from microphones and omni-directional vision sensor at the same time. The proposed system analyzes the relations of participants from capturer images and translates voice data to character data with speech-recognition technique. The proposed system classifies the translated character data into main speech and chattering by utilizing analyzed data, and displays character data of the main speech.

2. CONCEPTUAL DESIGN OF PROPOSED SYSTEM

2.1 Overview of proposed system

Our proposed meeting support system consists of four parts; image recognition part, speech recognition part, information integration part and display part. Figure.1 shows the conceptual design of proposed system.

©2006 ICDVRAT/University of Reading, UK; ISBN 07 049 98 65 3

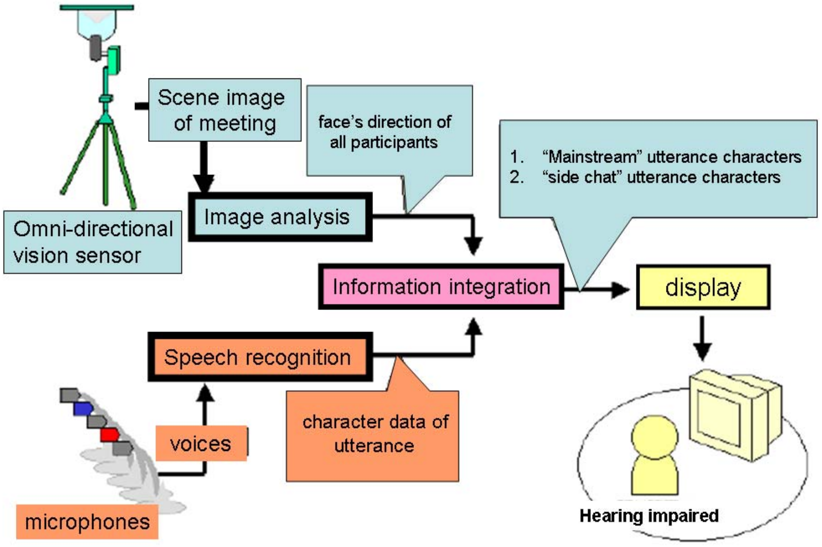

Scene image of meeting
face's direction of all participants
1. "Mainstream" utterance characters
2. "side chat" utterance characters
Omni-directional vision sensor
Image analysis
Information integration
display
Speech recognition
voices
character data of utterance
microphones
Hearing impaired

Figure 1. *Conceptual design of proposed system.*

First of all, the proposed system captures scene images of meeting from omni-directional vision sensor and captures voice data of all speakers from microphones. Secondly, the image analysis part in proposed system detects the face directions of all speakers from captured images and analyzes the group of speaking the mainstream of the discussion in the meeting. The speech recognition part translates captured voices to characters. Thirdly, the information integration part distinguishes characters of the "mainstream" and ones of "chat" by using the information from image analysis part and speech recognition part. Lastly, display part shows the mainstream of the discussion as characters. The proposed system would enable the hearing impaired to understand the keystone of the meeting they join.

2.2 Image processing of face direction detection

The mainstream of the discussion would be a set of important utterances in the meeting. Therefore, a speaker, who told an opinion related the mainstream, is the attention at the meeting and other participants of the meeting would watch the speaker. Thus, the detection of face directions enables the proposed system to find the mainstream of meeting. The image recognition part recognizes face directions of meeting's participants from captured omni-directional images.

The proposed system utilizes the omni-directional vision sensor. This vision sensor can capture 360 degree's images. Comparing with using some cameras, the omni-directional vision sensor is not necessary to synchronize the cameras for detecting face directions. This vision sensor is put on the center of table in meeting space to capture face data of all participants. The proposed system estimates face directions in the following manners. First, the system makes a pseudo-panorama image from a captured omni-directional image (Figure.2). Second, the system extracts face areas from the panorama image. Third, the system detects the face direction by using two methods; face recognition with sobel filter and face recognition with color information (R Brunell et all 1993, Y.Dai et all 1995).

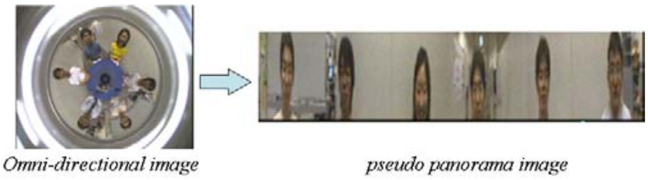


Figure 2. *Omni-directional image and panorama image.*

2.2.1 Extraction of face areas. The system makes one forth resolution image from the pseudo-panorama image from mosaic image process in order to run in a real-time and extracts the skin-color-areas and hair-color-areas from the pseudo-panorama image. Figure3 shows the process of face area extraction.

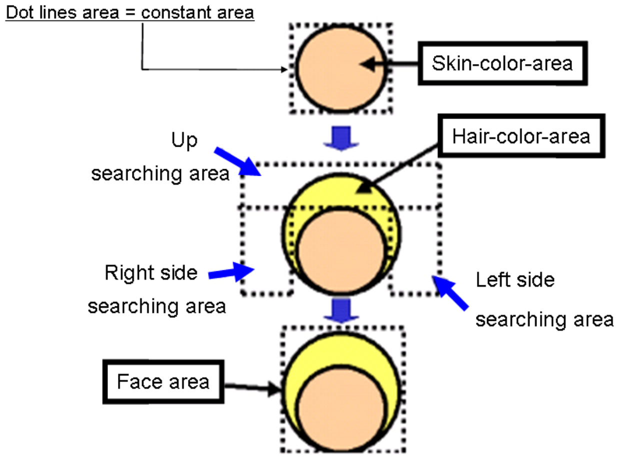


Figure 3. *Overview of face area extraction.*

To extract the skin-color area, firstly, each element in RGB colors is normalized. Next, when the normalized color is in the range of defined color parameters, the system defines the normalized colors as skin-color region. Thirdly, the skin-color regions are labeled. The system defines the labeled as the skin-color area, if the labeled has the constant area. To extract of the hair-color-area, RGB color is converted to Y/Cr/Cb color. If Y/Cr/Cb color is in the range of defined values, the Y/Cr/Cb is defined as the hair-color-area.

The system detects face area by using these color-areas; skin-color-areas and hair-color-areas. The system searches for the hair-color-areas near the skin-color-areas. If the hair area was found, the system defines both the hair-color-areas and skin-color-areas as the face areas.

2.2.2 Detection of face direction. As the method to detect face direction, the system utilizes two detection results of edge information and color information. Edge detection filter, Sobel Filter, is utilized to acquire the edge information. Sobel filter is one of the edge detection filter. When face image is processed by this filter, the outlines of mouth, nostril and eyes in one's face, a large gap of pixel values, are detected.

The proposed system makes a histogram by the acquired edge image. The histogram's horizontal axis is position and its vertical axis is a number of pixels. The histogram has the following features. When a speaker turns toward the front, the outlines of mouth, eyes are centered in the face area. Thus, the edge distribution's median is centered in the histogram. When a speaker turns to the right side, the outlines of face parts are right-sided in face area and the distributions median is also right side. When a speaker turns to left side, the edge distributions median is also left-side. Therefore, the proposed system detects the face direction, front, right and left from using the edge distribution median.

The proposed system calculates the median-value of edge histogram and distinguishes three directions of participant's face. In addition, the proposed system makes the histogram of skin-color area, and it also calculates the median value of the skin-color-histogram. The system recognizes the face direction by using the relation of these median values Figure.4 shows the processing of face direction detection.

If it is difficult to detect the face direction by using the differences between the median value of edge and one of skin-color, the system uses the following processing to detect face direction. The processing is to compare the median value of skin-color-area with the median value of hair-color-area. Both median value of skin color and median value of hair color depend on the face direction. Therefore, it can detect one's face direction to compare these median values.

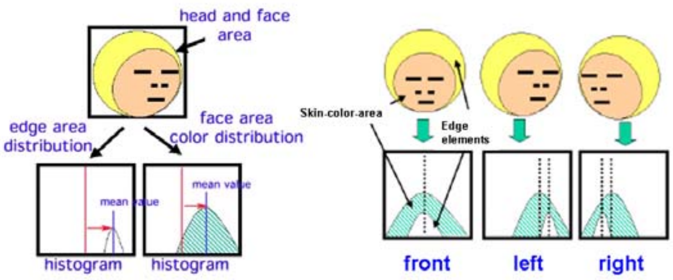


Figure 4. *Face direction detection.*

2.3 Speech recognition Process

It is necessary to display the contents of meeting so as to let the hearing impaired join in the meeting. The speech recognition part translates voice data to characters automatically. Output data from the speech recognition part is character data corresponded to speakers' voices. A number of microphones is equal to one of participants in meeting. To make the prototype system, IBM ViaVoice software is utilized in this paper.

2.4 Information Integration Process

The information integration part, one of the parts in the component of proposed system, classified the speeches of meeting into two groups; a group of "mainstream" and a group of "chat". This part performs the analysis of grouping from both output data of image recognition part and one of speech recognition part

The following procedures are processed in this part. Firstly, it is determined the groups the speaker belongs to from the relation of next participants.Figure.5 shows the relation of next participants to make the groups.

The rules to make groups are shown the below.

[RULE 1] When the proposed system recognized that a speaker and next speaker faced each other from the face direction, they belong to the same group.

[RULE 2] When the proposed system recognized that a speaker and next speaker looked in the opposite direction, they do not belong to the same group, but belong to the difference group each other.

[RULE 3] When it is not included in the RULE1 and RULE2, they do not have the group to belong.

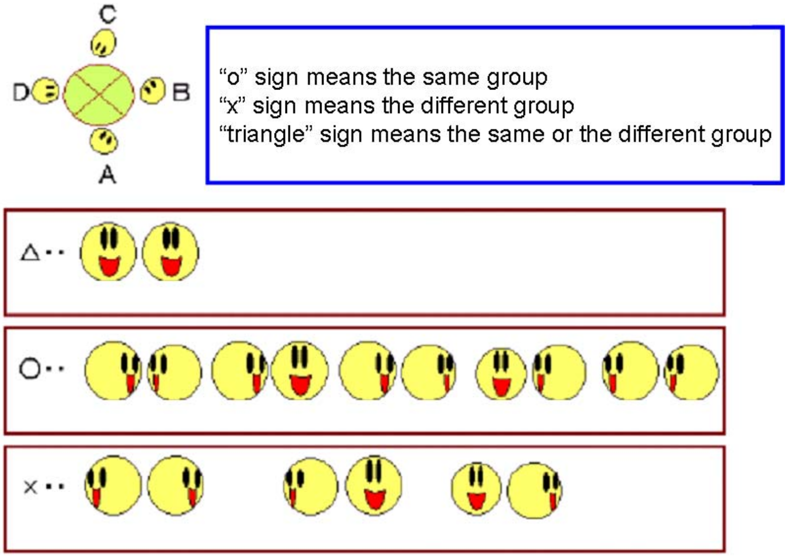

C
D
B
A
"o" sign means the same group
"x" sign means the different group
"triangle" sign means the same or the different group
△··
O··
×··

Figure 5. *Rule to make a group between next speakers.*

Secondly, this part makes two groups of "mainstream" and "chat" from groups made under the above rules.

The boundary of a difference group determined by RULE2 is regarded as the boundary of a large group. This part makes large groups from the groups made by RULE1-2. Then, when a group is made by the participants included n RULE 3, the group is one of elements in "mainstream" group or "chat" group. Figure 6 shows the case of groups. This part recognizes the group included in largest number of participants as the "mainstream" group. If some groups exist, which have the same participants, this part recognizes the group the speaker is belonged as the "mainstream" group.

2.5 Display

The display part shows the participant's name, conversation in each group on the screen. The contest of discussion is displayed in a box called "discussion box". The boxes of all groups are drawn.

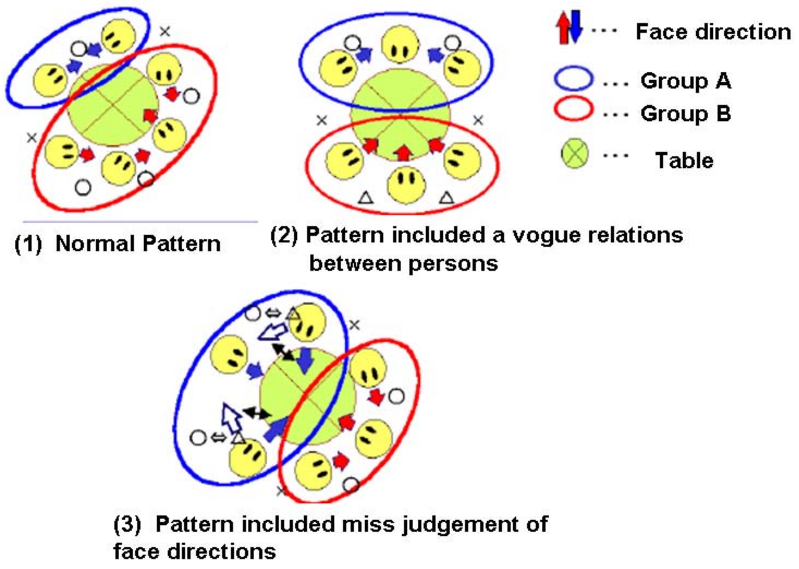

Face direction
Group A
Group B
Table
(1) Normal Pattern
(2) Pattern included a vogue relations between persons
(3) Pattern included miss judgement of face directions

Figure 6. *Overview of some grouped situation.*

3. EXPERIMENT

This paper has an experiment to evaluate the recognition rate of face direction method.

As the experimental plan, the meeting was ready for this experiment. The participants of the meeting are 6 persons. As the situation of the meeting, 4 case studies were set up in this experiment. These case studies are shown in Figure 7.

Case (A) indicates that all participants look at the front. Case (B) shows all participants look at a next person of right side. Case (C) shows that participant of label 1 looks to the right side, participant of label 2 looks straight, participant of label 3 looks to the left side, participant of label 4 looks to the right side, participant of label 5 looks to the left side and participant of label 6 looks to the right side. Case (D) is situation that participants of label 1, 2, 3, 5 look at the front, participant of label 4 looks to the right side and participant of label 6 looks to the left side. 1,000 images are captured in each case. The face directions are extracted by using the captured images.

Image resolution, utilized by the proposed system, is 330 x 240 pixels, and the size of face areas in captured image is about 30 x 60 pixels.

3.1 Experimental Result and consideration

Figure 8 shows the face direction results of each case. The horizontal axis is the change of face direction. The vertical axis is a number of images. The proposed system recognized the face direction as the right side, when the change value is larger than 2. The system recognized the face direction as left side, when the change value is smaller than -2. The proposed system recognized the face direction as the front, when the change value is between -2 and 2.

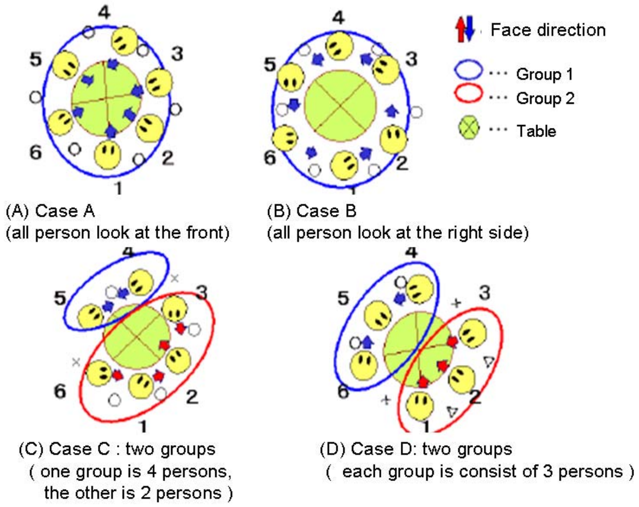
In Case A, the graph of Case A shows that the peak of distribution of the change is zero value in all 6 persons. Therefore, the proposed system recognized the face direction as the front and the system could recognize correctly the situation of Case A in this experiment.

In Case B, the graph of Case B shows that the peaks of distribution of all subjects are larger than 2.Therefore, the result showed that the proposed system recognized that all subjects turn to right side and recognized the situation of Case B in this experiment.

In Case C and Case D, these graphs showed that the proposed system recognized the each situation.

From the results of these cases, the recognition rate of face direction was about 90 percents.

In addition, the proposed system was classified the groups by using these recognition results. As the results, in Case A, the proposed system recognized 6 subjects as one group under the recognition rate of 75 percents. In Case B, the recognition rate was 30 percents. The classification of group was low recognition rates. The face direction by using edge detection was good results to judge the direction of each person, but because the small changes in edge distributions were influenced on the classification of groups, the proposed system would be difficult to recognize the groups.



(A) Case A
(all person look at the front)

(B) Case B
(all person look at the right side)

(C) Case C : two groups
(one group is 4 persons,
the other is 2 persons)

(D) Case D: two groups
(each group is consist of 3 persons)

Figure 7. *Situation of Experiment.*

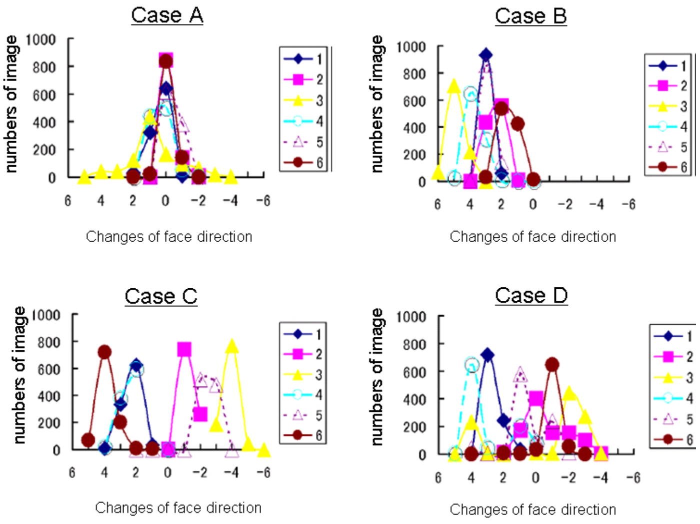


Figure 8. *Graphs of the change of face direction in each case.*

4. CONCLUSION

This paper describes a new meeting support system for the hearing impaired. The proposed system use edge distribution and skin-color-distribution to recognize a group of the mainstream of the discussion. The recognition method was evaluated in one experiment, the result showed effectiveness.

5. REFERENCES

R. Brunelli and T. Poggio(1993), Face Recognition: Features versus Templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15,No.10, pp.1042-1052

Y. Dai, and Y. Nakano(1995), Extraction of Facial Images from Complex Background Using Color Information and SGLD Matrices, Proc. Intl. Workshop on Automatic Face and Gesture Recognition, pp.238-242

©2006 ICDVRAT/University of Reading, UK; ISBN 07 049 98 65 3