Journal of the Korean Institute of Illuminating and Electrical Installation Engineers Vol. 17, No.5, pp. 68~79 September 2003 논문 17-5-10

영상회의를 위한 화자 검출 시스템

(Speaker Detection System for Video Conference)

이병선* · 고성원 · 권혁봉
(Byung-Sun Lee · Sung-Won Ko · Heak-Bong Kwon)

요 약

본 논문에서는 여러 사람이 참여하는 영상 회의에서 입술 움직임 정보를 이용하여 화자를 검출하는 시스템을 구현하였다. 구현된 시스템은 얼굴색 정보와 형태 정보를 이용하여 각 사람의 얼굴 및 입술 영역을 검출한 후, 입술 영역에서 이전 프레임과의 변화량을 계산하여 화자를 검출한다. 검출된 화자를 클로즈업하기 위하여 두 대의 CCD 카메라를 사용하였으며, RS-232C 시리얼 포트를 이용하여 PTZ 카메라를 제어한다. 실험 결과 3인 이상의 입력 동영상에서 얼굴의 기울어짐에 무관하게 화자를 검출할 수 있었으며 최초 기준 영상에서 화자를 클로즈업하는데 약 4~5초 정도의 시간이 소요되었다. 또한 320×240 크기의 얼굴 영역 화면과 전체적인 배경 화면을 동시에 제공하므로 영상회의 및 인터넷 방송 등과 같은 영상 전송 시스템에서 보다 효율적인 의사 전달이 가능하게 하였다.

Abstract

In this paper, we propose a system that detects the current speaker in multi-speaker video conference by using lip motion. First, the system detects the face and lip area of each of the speakers using face color and shape information. Then, to detect the current speaker, it calculates the change between the current frame and the previous frame. To accomplish this, we used two CCD cameras. One is a general CCD camera, the other is a PTZ camera controlled by RS-232C serial port. The result is a system capable of detecting the face of current speaker in a video feed with more than three people, regardless of orientation of the faces. With this system, it only takes 4 to 5 seconds to zoom in on the speaker from the initial image. Also, it is amore efficient image transmission system for such things as video conference and internet broadcasting because it offers a face area screen at a resolution of 320×240, while at the same time providing a whole background screen.

Key Words : face area detection, lip motion, speaker detection, PTZ camera

1. Introduction

The rapid development of remote video conference technology has been fostered by the advancement of information communication network system technology and image processing technology.

Remote video conference systems are now used in several fields, such as remote education, remote

* 주저자 : 김포대학 전자정보계열 조교수
Tel : 031-999-4169, Fax : 031-999-4775
E-mail : bslee@kimpo.ac.kr
접수일자 : 2003년 5월 27일
1차심사 : 2003년 6월 5일
심사완료 : 2003년 7월 18일

medical treatment, great distance watch, and international business consultation because there are no restriction of time or place and high capacity multimedia information, such as text messages, voice, and images can be transmitted in real time.

The speaker detection system is a system that recognizes the current speaker during video conference. As a lip movement method, Delmas and Coulon studied the speaker detection using helmet with camera[1]. But variables such as background, the distance between the speakers face and the camera, facial position in cameras visual field, were all obstacle to creating a system that detects the current speaker in real time.

The first thing needed to determine the current speaker is to detect the faces in the background image. To detect faces, aspects such as color, shape, statistic, and motion information are used. Sobottka and Pitas detected face using Hue and saturation, and Chai and Ngan made a skin color reference map from YCbCr color space which searched for face position via a normalization processes based region[2][3]. In addition to these, there are statistical methods such as PCA (Principle Component Analysis), and Maximum Likelihood[4][5].

The method that is used in this paper is based on color and morphological characteristics of face area. We separate face color area from Cb and Cr of Y, Cb, Cr images and make binary images. Binary images remove noise and small areas by using a morphological filter, and separate face candidate area via labeling process[6]. After the characteristic information of each candidate area is found, a search for eye area is conducted by applying the vertical operator of Sobel mask in Y image of face candidate area, and the jaw area is located by applying the horizontal operator of Sobel mask. Therefore, it detects exactly face area.

Among all the detected face areas, an algorithm, using lip motion information from real time changes in face area, is employed to detect the current speaker. First, we get the center of gravity of the face area in a binary image to look for lip area in detected face area. Lip area is easily separated using structural characteristics that are situated lower than the center of gravity of the face area. It detects the current speaker by calculating the amount change of "0" in lip area of binary image that changed according to change of lip shape.

2-channel image input devices and two color CCD cameras are used in order to zoom in on the detected speaker. The PTZ (Pan/Tilt/Zoom) driver is controlled with an RS-232C port. The position of detected speaker can be shown with the two-dimension coordinate of reference image, and then converted to the moving time of camera. It can move the focus of camera to the center of speaker, and zoom in on the speaker.

The remainder of the paper is organized as follows. Section 2 presents the algorithm of facial region detection used this paper. Section 3 describes the algorithm of speaker detection using lip reading. In section 4, we give examples of the experimental results and evaluate the performance of the system. The conclusion and possible future extension of the system are discussed in Section 5.

2. The face area detection algorithm

In order to take a close-up of the current speaker in real time from a video image that is taken from a basic camera, first all the face areas in that video image should be detected. In this paper, in order to detect the face area, we used color and morphological characteristics of face area. Figure 1 shows the process of the face area detection algorithm.

First, to decrease effect of lighting the RGB input image is changed to YCbCr color space. Then, the binary image is produced from the separated face color area that is removed from Cb and Cr, which are not effect by lighting very much. The Face color threshold proposed by Chai[7]. was showed in equation (1).

$$B(x,y)=\begin{cases}1 & \text{if}(77\le Cb\le 127)\cap(133\le C_r\le 173)\\ 0 & \textit{Otherwise}\end{cases} \quad (1)$$

B(x, y) is a binary image that is segmented by skin color. If Cb and Cr components are within the threshold, it regards that as skin color and sets to 1. Other parts are regarded as background image and set to 0. The image needs to be simplified by removing small noise areas through morphological filtering and overhang because areas with excess noise are included in the binary image that was acquired by color skin division with the threshold of Cb and Cr components.

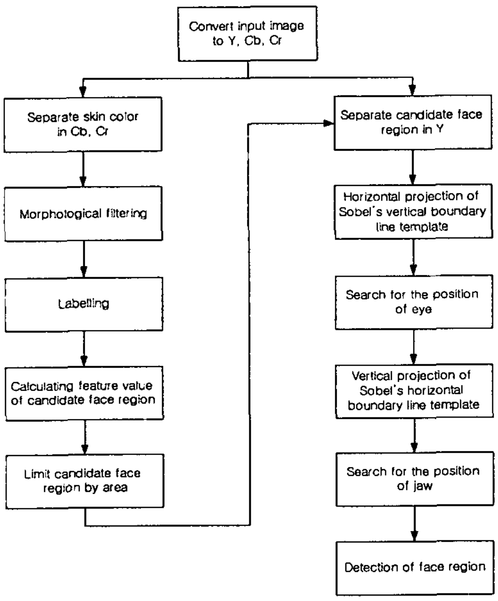


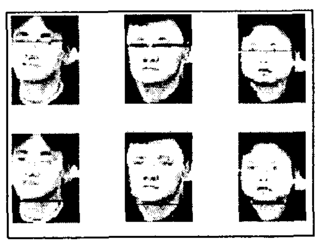
Fig. 1. The process of the face area detection algorithm

It separates each candidate face area using labeling techniques in the morphological filtered image and detects face area by limiting both the area and the ratio of width and length. By detecting the eye and jaw position within detected face area, it can set the lip area. First, it separates face area and detects the position of eye and jaw in Y component of each face area. To find the eye position in each face area, it uses Sobel's vertical boundary line template, and compares the position with center coordinate of detected face area.

If it is situated lower than the center coordinate, the lowest coordinate of detected face area is set to jaw position. If the eye position is situated above the center coordinate, it regards the neck part as belonging to the face area.

(a) Input image

(b) The detected part of face area



(c) The detected eye and jaw position

(d) The final result of face area detection

Fig. 2. The detected face area

To remove the neck part from the face area, it adds the routines that search for the jaw position. The position of the jaw can be detected using Sobel's horizontal boundary line template. To decrease the effect of noise, the position of jaw can be set the position that the values of horizontal projection of Sobel changes sharply from 5 pixel upsides in lowest position of face region.

Figure 2 expresses the detected face area. Picture (a) is input image; (b) is detected part of the face area by Cb and Cr component; (c) is the image of the detected eye and jaw position in the Y component of each face area; (d) is the final result of face area detection.

3. Current speaker detection algorithm

Like lip motion, to detect the movement of the object in the special part within the image, the speaker detection extracts characteristic values from the each image in the real time input image, and analyzes the change pattern of characteristic value between the continuous frames. In this case, the methods based on pixel value or histograms are mainly used[8][9]. In this paper, the amount of lip motion change is used to detect the speaker, and is based on a histogram method, but does not lose position information. After setting the proper lip area in the detected face area, it extracts motion information by measuring the pixel area changes in the lip area of the binary image that is separated by face color.

3.1 The detection of lip area

The proposed speaker detection system in this paper detects lip motion rapidly at the expense of absolute correct detection. Therefore, it sets lip area from the jaw line upward, and in order to calculate lip motion information quickly, only the change in pixel values of lip area are calculated. Through the results of the morphological analysis of facial image[10] and an experiment, it was determined that the most suitable size for the lip area mask is 19x19 pixel when size of face area is set to 700~1500 pixels in a 320240 image.

3.2 The detection of lip motion information

In this paper, to detect the current speaker in the video image acquired from the CCD camera, it extracted the change in area of "0" within the lip area mask between two frames of (n-1) and n, and defines the amount of change to this areas as lip motion information. When the speaker talks, the face color is not changed, but rather the number of the pixels in the lip area mask of the binary image are changed, which are segmented as skin color by inside lip shade and teeth. In this study, we discovered that lip motion information was acquired using the change in area of pixel within this lip area mask.

In order to find the lip motion information, we detect the face area every 15 frames and then calculate the center coordinate and reset lip area mask. We use equation (2) to detect the amount of change of lip motion information between two comparison frames for each person.

$$\begin{aligned}\delta_n &= \mid S_1 - S_2 \mid \\ &= \mid \frac{1}{n}\sum_{i=1}^{19}\sum_{j=1}^{19} f_1(x_i, y_j) \\ &\quad - \frac{1}{n}\sum_{i=1}^{19}\sum_{j=1}^{19} f_2(x_i, y_j) \mid \end{aligned} \tag{2}$$

In this case, fn (xi, yi) if pixel value of (xi, yi) position in lip area mask is "0", the value is 1, if not the value is 0. Sn is area of 0 in lip area mask for each frame area, and δn displays motion information of lip by the amount of change in area in mask between two frames.

3.3 The detection of the speaker

In order to detect the current speaker, we detect face area from an extract of the first frame of the video image taken from the color CCD camera for the candidate speakers using face color information and morphological information.

Also, we established the lip area mask for each candidate speakers' face area and calculated the area of "0". Then, each candidate speakers' maximum lip motion information amount is calculated. Following these steps leads to the detection of the current speaker.

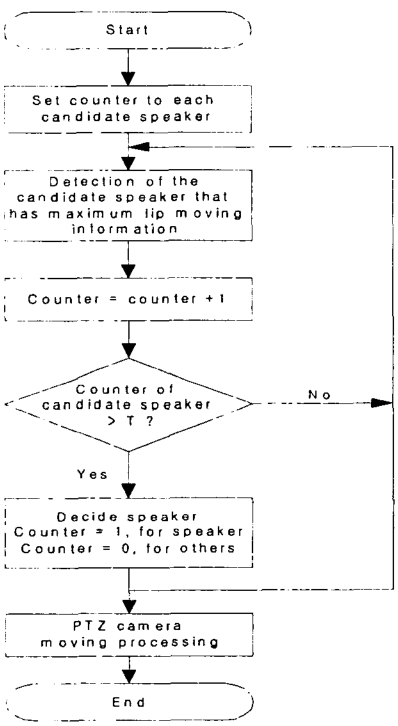


Fig. 3. The flowchart of the speaker detection algorithm

Figure 3 shows the flowchart of the speaker detection algorithm. "T" is threshold value for lip motion detection and was determined to "5" by experiment. To calculate the amount of maximum lip motion information, we searched the changed binary image every 15 frames, and extracted, in real time, the lip motion information of all candidate speakers from the input image. Figure 4 expresses the process of extracting the lip motion information of all candidate speakers in the video image.

If each candidate speakers are labeled A, B, C, the area of first area mask for each candidate speaker is expressed as mask A0, mask B0, mask C0, and the area of area mask after 15 frames is by mask A1, mask B1, mask C1.

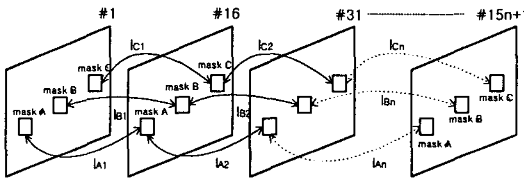


Fig. 4. Calculation of lip motion for each frames

If lip motion information between frame number 1 of each candidate speaker and frame number 16 is defined as IA1, IB1, IC1, between frame number 16 and frame number 31 is defined as IA2, IB2, IC2, and the continuous lip motion information of each candidate speaker can be defined as IAn, IBn, ICn. Therefore, lip motion information for each candidate speaker A, B, C can be shown as in equation (3).

$$\begin{aligned} I_{An} &= \mid (mask\ A_{n-1}) - (mask\ A_n) \mid \\ I_{Bn} &= \mid (mask\ B_{n-1}) - (mask\ B_n) \mid \\ I_{Cn} &= \mid (mask\ C_{n-1}) - (mask\ C_n) \mid \end{aligned} \quad (3)$$

Here, n is integer, and (n-1) and n express a 15 frame interval. Also, maximum lip motion information is the greatest value among the calculated IAn, IBn, ICn in equation (3). Figure 4 shows that we calculated this between beginning and end of each frame interval in real time with algebraic subtraction.

To choose the current speaker, we continuously compared the amount of change in lip motion of each candidate speaker. Through an experiment, TI is the threshold and set to 5. The candidate speaker with a maximum amount occurring 5 times is defined as the current speaker. Then, the number of the maximum change value of all candidate speakers, except the current speaker, are reset to "0", and the current speaker is set to "1".

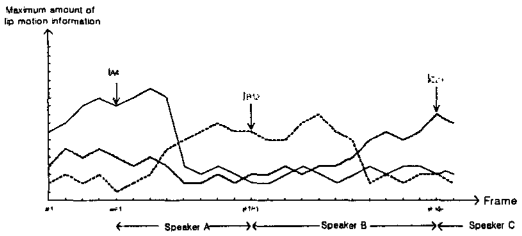


Fig. 5. Speaker decision for lip motion change of a candidate speaker

This is done to ensure the speakers chosen consistently, even with a momentary change in maximum lip motion information in the candidate speakers. Figure 5 shows a graph of the process of selecting the speaker. In this graph, the lip motion information amount of the candidate speakers was decided arbitrarily, and the candidate speaker whose maximum change amount occurs 5 times is speaker.

3.4 PTZ camera controlling

To take a close up of the speaker candidate, whom the detection algorithm determined was the current speaker among the candidate speakers, we use another CCD color camera that possess PTZ function, and control the motion of camera using RS-232C serial port. First, figure 6 expresses the standard image at a size of 320×240, which is used in the face image signal processing, as a two dimensional coordinate system, and calculates the coordinate CA (Xn, Ym) that correspond to center of current speaker's face area. Because the PTZ camera movies up, down, left, and right in set time intervals, to move the area expressed by the two dimensions coordinate, the camera moves by converting the function of moving time.

In figure 6, the central point of the image is defined as 0(X0, Y0), and assuming an arbitrary A point as the current speaker's central point, the distance between point A (Xn, Ym) and the X axis is separated by n pixel, and the Y axis by m pixel from the standard point 0(X0, Y0). Therefore, to move the center of camera to point A, the distance of pixel for X, Y coordinate has to be converted to a camera moving time unit. To move the Pan/Tilt precisely we conducted the experiment repeatedly.

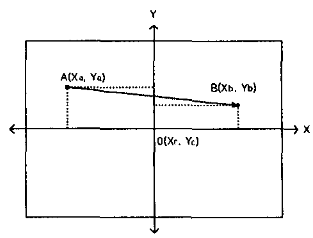


Fig. 6. The coordinate system for reference pictures

We found that the camera focus, to move one pixel, takes 15ms' moving time. Equation (4) expresses the moving time of camera depending on the distance on the two-dimensional coordinate system.

$$X_n(t) = n \times 15[\text{ms}]$$
$$Y_m(t) = m \times 15[\text{ms}] \quad (4)$$

Also, to minimize moving distance of camera focus when the speaker changes, the camera moves directly to the next speaker's central point on the basis of the current speaker. Thus, if an arbitrary speaker A is decided, the camera moves to speaker A, and the standard point is then changed from central point 0 (X0, Y0) of coordinate system to speaker A's central point A (Xa, Ya). Therefore, in case where the speaker becomes person B (Xb, Yb) as in figure 6, the moving distance of camera is calculated with equation (5).

$$X_M = |Xa - Xb|$$
$$Y_M = |Ya - Yb| \quad (5)$$

4. Simulation result

In this research, we designed and established a system that can detect and take a close-up of the speaker's face area in the video image that are taken from a CCD color camera. To construct the system, it detects the face area using color information, extracts constantly changing lip motion information and detects speaker in real time. A close up of the speaker's face area was taken PTZ camera. By showing both background image and speaker's face area in one image, we gain two pieces of information from two images.

4.1 System configuration

The experiments were performed on an Athlon 1Ghz PC with an algorithm that was implemented using Visual C++ 6.0. We used two cameras a SONY TRV900 as a reference camera, a 128 times zoom camera with PTZ function. We used RS-232C port for communication between the PTZ camera and the PC. The camera control was done by the PTZ camera receiver. In order to input image signal from two cameras, we acquired a 30frame/s, 320×240 size video image using a 4-channel DVR card, Eyean-1000.

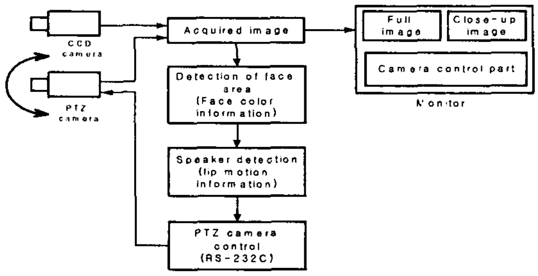


Fig. 7. Hardware configuration for speaker close-up system

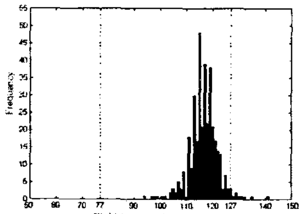
Figure 7 shows the hardware configurations for speaker close-up system. The speaker close-up system, which we constructed to detect the current speaker, detected face areas of all candidate speakers in video image that was taken from the CCD color camera in intervals of 15 frames.

4.2 Simulation of lip motion information detection

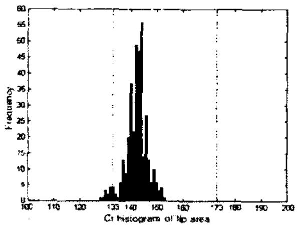
We set the lip area for the people's faces within video image, and analyze the Cb and Cr histogram and the shade histogram for lip area regardless of whether a candidate is speaking or not. We calculate area change of "0" in the binary image based on the face color change of the lip area.

Fig. 8. Face image for three candidate speakers (first speaker)

Also, by calculating "0" area difference of set lip area, we express the degree of lip movement as the number for "0" pixel. If the number for "0" pixel is continuously changing, it means the lips are moving, and therefore main factor for deciding the current speaker.

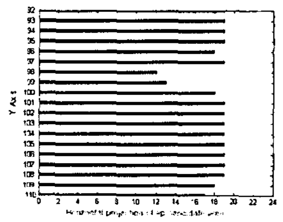


(a) The histogram of Cb components

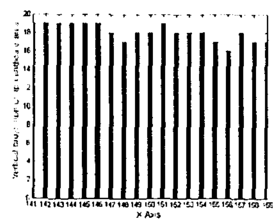


(b) The histogram of Cr components

Fig. 9 The histograms lip area



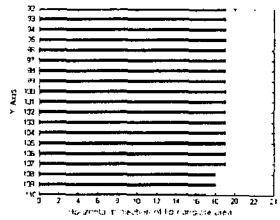
(a) Horizontal projection



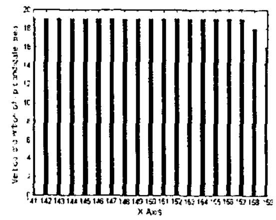
(b) Vertical projection

Fig. 10. Projection histograms of (n-1)-th frame

Figure 8 shows image when the first candidate speaker, among three, is the current speaker. Figure 9 shows the histogram of Cb, Cr within the lip area mask when the first speaker talks. We can see that Figure 9 is beyond the extent of the threshold of face color, as shown in equation (1).



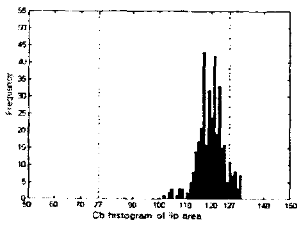
(a) Horizontal projection



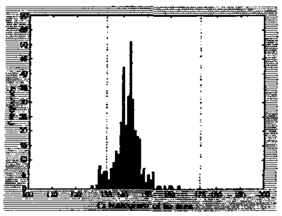
(b) Vertical projection

Fig. 11. Projection histograms of n-th frame

Fig. 12. Face image for three candidate speaker (second speaker)

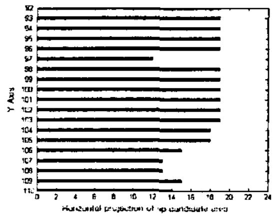


(a) The histogram of Cb components

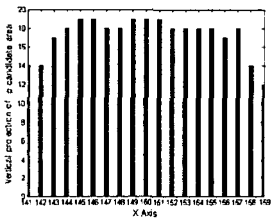


(b) The histogram of Cr components

Fig. 13. The histograms of lip area when speaking

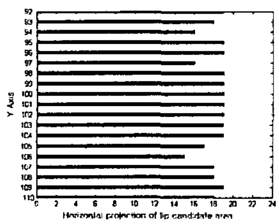


(a) Horizontal projection

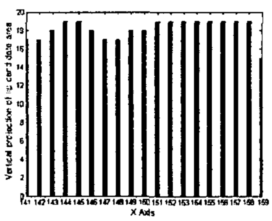


(b) Vertical projection

Fig. 14 Projection histograms when speaking



(a) Horizontal projection



(b) Vertical projection

Fig. 15. Projection histograms when not speaking

Figure 10 and 11 show horizontal and vertical projections of the mask area for n-th frame and (n-1)-th frame.

The purpose of these is to show the change of "0" area within the lip area mask. In order to make a comparison with the cases in figures 8~11, figure 12~15 illustrate the change of lip movement between the first speaker and the second speaker.

Table 1 shows the lip motion information of each candidate speaker, found in figure 8 and 12, as the "0" number. Lip motion information (1) displays the difference of "0" number between the first frame and the second frame, and lip motion information (2) displays the difference between the second frame and the third frame.

Table 1. The calculations of lip motion information for each frame

"0" numb	1'st frame	2'nd frame	3'rd frame	Lip motion information (1)	Lip motion information (2)
Speaker A	18	9	5	9	4
Speaker B	30	36	16	6	20
Speaker C	9	16	18	7	2

For performance comparison, we compared our proposed lip motion detection algorithm to binary image difference method, which is the general method of motion detection. The binary image difference method achieves the image difference by calculating the difference between previous image and the current image. In this way, movement can be detected. Experiments show that lip movement is equally detected in both the proposed method and the binary image difference method when the speaker does not move very much.

However, when a speaker candidate, while not speaking, moves, such as rotating or lowering their face, the binary image difference method will make errors in detecting the lip motion. On the other hand, the proposed lip motion detection algorithm method detects lip area in face area even when the face moves.

Table 2. The comparison between difference image method and proposed method for each frame

Methode	Speaker (A)	Speaker (B)	Speaker (C)
Difference image	23	38	27
Proposed methode	9	6	1

Table 2 shows that the proposed method extracts the information of lip movement even if face moves. In this case, the actually speaker was speaker (A), however the binary image difference method detected speaker (B), while the proposed method detect speaker (A).

4.3 Camera control

Figure 16 shows the center point of three candidate speakers, and Figure 17 shows the central position of each candidate speaker on a two-dimensional coordinate system. The central position of each candidate speaker are A(-104,13), B(1,31), C(85,30). It takes 1.77 seconds, 0.48 seconds, 1.72 seconds respectively for the camera focus to move from 0(X0, Y0) to the center of each speaker's position. Table 3 shows the distances among candidate speakers and the moving time when the speakers change.

Fig. 16. Centroid of three candidate speakers

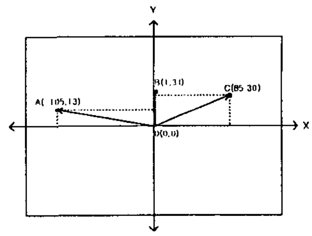


Fig. 17. The 2-D coordinate system for three candidate speakers

Table 3 Coordinate value and camera moving time

Speaker	Coordinate system	Moving time at 0 (0, 0)	Moving time (A to B)	Moving time (A to C)
Speaker A	(-105,13)	1.77 sec.	1.86 sec.	3.1 sec.
Speaker B	(1,31)	0.48 sec.		
Speaker C	(85,30)	1.72 sec.		

4.4 Current speaker detection simulation

In order to find the current speaker among the candidate speaker, we calculate the difference in area of "0" every 15 frames of video image input, and define this as lip motion information.

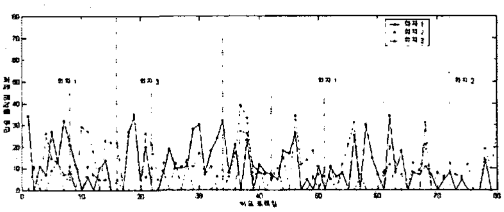


Fig. 18. Speaker decision process for three candidate speakers

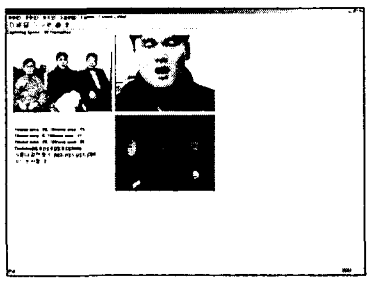


Fig. 19. Simulation result of the proposed system

Also, if a candidate has the maximum value more than 5 times, we choose that speaker as the current speaker. Figure 18 shows a graph of the maximum information of lip motion for a 1,201 frame interval for three candidate speakers. It takes about 4.5 seconds, occurring at the 136-th frame, to determine the first speaker among three candidate speakers. Then, when speaker changes, it takes 4 to 5 seconds to zoom in on the new speaker.

We created a simulator for performance estimation of the proposed real-time speaker zoom in system. Figure 19 shows the final image. The top left hand side is a picture of the three candidate speakers, and the candidate face area by color is shown on bottom right hand figure. The speaker on the top right hand side is the current speaker as chosen by the speaker detection system through lip motion and the bottom left hand side shows the information.

5. Conclusions

In this paper, we proposed a face area and speaker detection algorithm through lip motion, designed a system that can zoom in on the speaker and greatly magnify the face of the current speaker among all the candidate speakers.

The algorithm of face area detection used face color information and morphological information for real time processing, and was divided into face color area using Cb and Cr components from color space of YCbCr to decrease effect of lighting. We used morphological filtering and geometrical fitting for noise exclusion and simplification of the area of the candidate's face was achieved by segmenting face color. The filtered image was segmented into individual candidate face areas through a labeling processing. In addition, to detecting face area, the proposed system used Sobel's transformation to limit the ratio of width and length, and search for the position of eyes and jaw.

In order to detect the current speaker from the detected face areas of the candidate speakers, we calculate the amount of change of "0" in the lip area, which changes according to change of speaker's lip shape. The results of our experiment showed the mean value of information of actual speaker's lip motion was approximately 14 pixel, not speak was 7 pixel.

To avoid choosing the incorrect speaker, the algorithm of speaker detection chose a candidate speaker who maximum information of lip motion occurred 5 times. In experiments conducted in a laboratory environment, the current speaker could be detected within approximately 2.5 seconds in video image that has 3 candidate speakers.

We used a color CCD camera with a built in PTZ function to take a close-up of the speaker who is detected by proposed algorithm. The camera was controlled using a RS-232C serial port. To move the focus of camera to the central point of each speaker, the standard image was transformed to a two-dimensional coordinate system, and then the distance of speakers was converted to the moving time of camera and controlled camera.

The moving time of camera was a rapid speed of 15ms/pixel, and it took an average time of 2 seconds to move from standard image to current speaker's position. Therefore, it takes about 4 to 5 seconds to zoom in on the current speaker in the close-up speaker system. Also, because the proposed system was designed so that the focus of the camera follows the central point of speaker's face, it's application can magnify such as object tracing system.

Because the speaker detection system was constructed using the amount of change in lip motion, it provides a face area screen of 320x240 size and whole background screen at the same time. Because this system very efficiently transmits speaker's image through image transmission system such as video conference and Internet broadcast, we believe that this system can have many practical applications, including but not limited to remote watch systems, great distance object identification systems, and object tracing systems.

이 논문은 2003학년도 김포대학의 연구비 지원에 의하여 연구되었음.

References

〔1〕 P. Delmas, P. Y Coulon, and V. Fristot, "Automatic Snakes for Robust Lip Boundaries Extraction", 1999 IEEE International Conf, Vol.6, Acoustics, Speech and Signal processing pp.3069-3072, 1999.

〔2〕 K. Sobottka and I. Pitas, "Extraction of Facial Regions and Features using Color and Shape Information," IEEE Proc. Pattern Recognition, vol. III, pp.421-425, 1996.

〔3〕 D. Chai and K. N. Ngan, "Location facial region of a head-and-shoulders color image," IEEE Proc. Automatic Face and Gesture Recognition, pp.124-129, 1998.

〔4〕 M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," IEEE Proc. Computer Vision and Pattern Recognition, pp.586-591, 1991.

〔5〕 A. J. Colmenzarez and T. S. Huang, "Maximum Likelihood Face Detection," IEEE Proc. Automatic Face and Gesture Recognition, pp.307-311, 1996.

〔6〕 Young-Gil Kim, Jae-Hyeok Han, and Jae-Hyeong Ahn , "Facial regions detection using the color and shape information in color still images", Journal of korea multimedia society, Vol. 4, No. 1, pp.67-74, 2001.

〔7〕 D. Chai and K. N. Ngan, "Locating facial region of acolor image," IEEE Proc. Automatic Face andRecognition, pp.124-129, 1998.

〔8〕 S. Nagaya, T. Miyatake, T. Fujita, W. Ito, and H. Ueda, MovingObject Detection by Time Correlation Based BackgroundProceeding of ACCV '95, pp.717-722, 1995.

〔9〕 Ok-Sam Chae, Jeong-Heon Lee, Yong-Hak Ahn, and Seong-Guk Lee, "Invader watching and tracing system using neural network" The 8th image processing workshop, pp.167-172, 1996.

〔10〕 Tae-Ung Yu, and Oil-Seok Oh, "Facial region detection based color distribution information", Korea information science society, Vol. 24, No. 2, pp.180-192, 1997.

◇ 저자소개 ◇

이병선 (李秉善)

1958년 8월 5일생. 1985년 2월 서울산업대학 전자 공학사. 1992년 3월 건국대학교 대학원 공학석사. 2000년 2월 단국대학교 대학원 공학박사. 1994-1997 경북전문대학 전자과 전임강사. 현재 김포대학 전자정보계열 조교수.

고성원 (高成元)

1960년 8월 6일생. 1983년 2월 한양대학교 전자 공학사. 1995년 2월 한양대학교 대학원 공학석사. 1986-1995 한국통신 통신망 연구소. 1995-1997 영월전문대학교 전자과 전임강사. 현재 김포대학 전자정보계열 조교수.

권혁봉 (權赫奉)

1964년 2월 27일생. 1990년 2월 호서대학교 정보통신공학과 공학사. 1992년 2월 호서대학교 대학원 공학석사. 2002년 충북대학교 대학원 공학박사. 현재 김포대학 전자정보계열 조교수.