스테레오 시청각 기반의 화자 검출 시스템☆

A Speaker Detection System based on Stereo Vision and Audio

안 준 호* 홍 광 석**
Jun-ho An Kwang-Seok Hong

요 약

본 논문에서 다수의 사용자 중에서 현재 발성하고 있는 화자를 검출하는 스테레오 시청각 기반의 화자 검출 시스템을 제안한다. 제안한 시스템은 두 개의 마이크를 이용한 음원 위치추정, 스테레오 카메라를 이용한 영상정합 및 발화자 후보 위치 추정, 그리고 모바일 기반의 화자 검출 정보 획득으로 구성되어 있다. 스테레오 카메라로부터 획득한 화자의 영상정보를 바탕으로 Adaboost 알고리즘과 Haar-like 특징을 이용하여 발화자 후보들의 얼굴을 검출하고 이를 기반으로 삼각측량법을 이용하여 발화자 후보들의 위치를 추정한다. 그리고 2개의 마이크로부터 획득한 화자의 음성정보를 바탕으로 CPSP(Cross Power Spectrum Phase)기반의 TDOA(Time Differnce of Arrival)추정을 통해 음원의 방향을 추정한다. 최종적으로 스테레오 카메라를 통해 측정된 정보와 마이크를 통해 얻은 정보를 비교 분석하여 현재 발화자를 검출한다. 검출된 화자 정보에 대한 보다 차별화된 서비스 제공을 위해 TCP 서버/클라이언트 구조 기반의 모바일 화자 검출 정보 획득 시스템을 구현하고 평가하였다.

ABSTRACT

In this paper, we propose the system which detects the speaker, who is speaking currently, among a number of users. A proposed speaker detection system based on stereo vision and audio is mainly composed of the followings: a position estimation of speaker candidates using stereo camara and microphone, a current speaker detection, and a speaker information acquisition based on a mobile device. We use the haar-like features and the adaboost algorithm to detect the faces of speaker candidates with stereo camera, and the position of speaker candidates is estimated by a triangulation method. Next, the Time Delay Of Arrival (TDOA) is estimated by the Cross Power Spectrum Phase(CPSP) analysis to find the direction of source with two microphone. Finally we acquire the information of the speaker including his position, voice, and face by comparing the information of the stereo camera with that of two microphone. Furthermore, the proposed system includes a TCP client/server connection method for mobile service

☞ keyword : 음원 위치 추적(Source Localization), 스테레오 비전(Stereo Vision), 화자 검출(Speaker Detection)

1. 서 론

최근 정보통신기술의 급격한 발전에 따라 인간의 능력 컴퓨터에 접목시키고자하는 HCI(Human Computer Interaction)분야에 대한 관심이 커져감에 따라 인간이 가지고 있는 능력을 모델링하기 위한 연구가 여러 분야에서 활발히 수행되고 있다. 그 중 대표적으로 영상정보를 이용하여 사람을 검출하고 추적하는 기술과 음성정보를 이용한 음원 위치 추정(Source Localization)에 대한 기술에 관심이 날로 커져가고 있다[1][2].

인간의 시각시스템을 컴퓨터로 모방하고 3차원 공간을 활용할 수 있는 스테레오 비전은 영상처리분야에 있어서 매우 중요한 부분으로써 현재 많은 연구가 진행되고 있고 물체와의 거리 및 각도를 측정하기 유용하기 때문에 산업용 로봇, 3D 영화 제작, 게임 제작 등에 성공적으로 적용되어

* 준 회 원 : 성균관대학교 휴대폰학과 석사과정
amadasv@skku.edu
** 정 회 원 : 성균관대학교 정보통신공학부 교수
kshong@skku.ac.kr
[2010/09/16 투고 - 2010/09/29 심사 - 2010/10/28 심사완료]
☆ 본 연구는 지식경제부 및 정보통신 연구 진흥원의 대학 IT 연구센터 지원사업(NIPA-2010-(C1090-1021-00 08)) 및 2010년도 한국연구재단의 지원을 받아 수행된 연구임.(s-2010-0055-000)
☆ 본 논문은 창립10주년 2010년도 한국인터넷정보학회 학술발표대회 최우수논문의 확정버전임.

져 왔으며, 활용 분야는 앞으로 더욱 광범위 할 것으로 예상되어 진다. 스테레오 비전은 3차원 공간상에 설치된 카메라의 기하학적 특성에 따라 좌, 우 카메라로부터 얻은 영상에서 상호간에 정합을 이루는 변이를 찾아내고 그 정보를 바탕으로 3차원 거리 정보를 추정하는 일련의 과정을 거친다.

음원 위치 추정기술은 마이크 배열에 입력되는 신호를 바탕으로 신호 간의 시간차를 구하는 TDOA(Time Delay Of Arrival)방법을 통해 음원의 방위각을 추정하고 발화자의 위치를 결정하는 기술이다[3][4][5]. 하지만 음원이라는 단일 정보만을 이용할 경우 잡음, 회전, 반사, 굴절 등에 의한 오차발생이 큰 문제점으로 지적되고 있고 이와 같은 음성 정보만을 사용한 화자 위치 추정 방법의 한계를 극복하기 위해 영상정보를 이용한 얼굴 인식 및 입술의 움직임 추적 등을 함께 사용하는 멀티 모달 위치 추정 시스템에 대한 연구가 최근 국내외로 많이 진행되고 있다[6][7][8].

본 연구에서는 기존에 연구되어온 사용자 인식을 바탕으로 스테레오 카메라와 마이크를 이용하여 발화자의 위치, 거리 및 얼굴 등의 정보를 검출하고 보다 차별화된 서비스 제공을 위해 이를 TCP기반의 서버/클라이언트 구조로 구성하여 모바일에서 화자 검출 정보를 확인할 수 있는 스테레오 시청각 기반 화자 검출 시스템을 제안하고 구현한다.

2. 관련 연구

화자 검출 추정 방법에는 크게 두 개의 마이크를 이용한 음원 위치 추정과 스테레오 카메라를 이용한 영상 정합 및 화자 위치 추정으로 구분된다. 음원 위치 추정 방법은 두 마이크에 전달된 신호를 바탕으로 주파수 도메인에서 CPSP(Cross Power Spectrum Phase)기법을 이용하여 TDOA (Time Delay Of Arrival)을 구한 후 음원의 방향각을 획득하는 방법을 사용하였다[3]. 이와 동시에 스테레오 카메라로부터 획득된 좌, 우 이미지에서 각각 Adaboost Algorithm과 Haar-like feature를 이용하여 얼굴을 검출한 후 이를 기반으로 영상 정합과 삼각 측량법을 이용하여 발화자의 거리와 위치, 얼굴 정보를 획득하는 방법을 사용하였다.

2.1 음원 위치 추정(Source Localization)

기존에 존재하는 음원 위치 추정 방법으로는 고해상도 스펙트럼 추정방법(spectral analysis), 빔 형성 방법(beamforming), 도착 시간 지연법(TDOA)등이 있다[2]. 하지만 무인 로봇이나 화상회의 시스템과 같이 마이크의 위치 및 간격이 플랫폼의 고정에 의해 제한되는 경우에는 빔 형성 방법이나 고해상도 스펙트럼 추정 방법은 적합하지 않다[4]. TDOA를 이용할 경우 다른 방법들보다 계산량이 적고 마이크의 배열이 상대적으로 자유롭지만 잡음에 상당한 영향을 받는 문제점이 있다. TDOA를 측정하기 위한 대표적인 방법으로는 주파수 도메인에서 두 신호의 상관성을 이용하는 CPSP기법이 있다. 음원에서 발생한 신호는 두 개의 마이크로 전달되는데 이 때 각각의 마이크에서 측정된 신호를 $x_1(t), x_2(t)$라고 정의하고 음원에서 발생한 신호를 $s(t)$, 노이즈를 $n(t)$라 정의하자. 수신신호는 식 1과 같이 표현할 수 있다.

$$\begin{aligned} x_1(t) &= s(t) + n_1(t) \\ x_2(t) &= \alpha s(t+D) + n_2(t) \end{aligned} \quad (1)$$

식 1에서 D는 구하고자 하는 두 신호사이의 시간지연이고 α는 감쇠상수를 나타낸다. 두 마이크로폰에서 받은 신호의 CC(Cross Correlation)은 다음과 같이 정의 된다.

$$R_{x_1x_2}(\tau) = E[x_1(t)x_2(t-\tau)] \quad (2)$$

식 2에서 $R_{x_1x_2}(\tau)$가 최대값을 가질 경우 D를

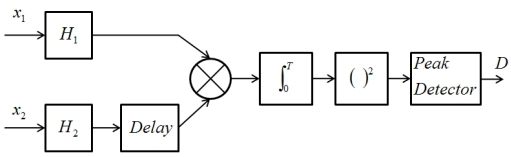
구할 수 있다.

$$D = \arg\max[R_{x_1x_2}(\tau)] \quad (3)$$

하지만 실제 환경에서 무한구간의 측정은 불가하므로 일정한 측정시간 T에 따른 수식으로 근사화 하게 된다.

$$\hat{R}_{x_1x_2}(\tau) = \frac{1}{T-\tau}\int_0^T x_1(t)x_2(t-\tau)dt \quad (4)$$

다음 그림 1은 CC방법을 이용한 피크검출의 흐름도이다.



(그림 1) Cross Correlation을 이용한 피크검출

여기서 상호 상관 함수$R_{x_1x_2}(\tau)$과 Cross Power Spectrum $G_{x_1x_2}(\tau)$는 다음과 같은 Fourier변환관계를 가진다.

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} G_{x_1x_2}(f)e^{j2\pi f\tau}df \quad (5)$$

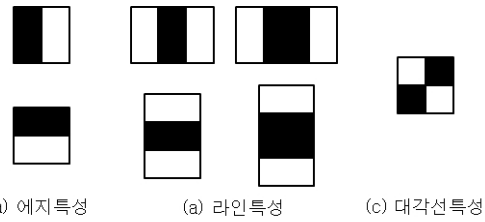
x_1과 x_2사이의 GCC(Generalized Cross Correlation)는 다음 식 6과 같이 식 5에 가중함수를 곱하는 형태로 정의된다. 가중함수의 종류에는 PHAT, SCOT등이 존재한다[5].

$$R^{(g)}_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \psi_g(f)G_{x_1x_2}(f)e^{j2\pi f\tau}df \quad (6)$$

수식에서 $\psi_g(f)$는 가중 함수로서 실제 구하고자 하는 지연시간의 값은 가중 Cross Power Spectrum 즉, $G^{(g)}_{x_1x_2}(\tau) = \psi_g(f)G_{x_1x_2}(f)$의 역 푸리에 변환 값이고 GCC함수를 최대화하는 값으로 구해진다. 여기서 가중함수의 크기를 1로 할 경우 Cross-Correlation 방법과 동일하다. 따라서 CPSP의 GCC함수의 특성은 Cross-Correlation의 최대값을 구할 때 용이하므로 그 성능이 뛰어난 것으로 알려져 있다.

2.2 스테레오 카메라를 이용한 발화자 후보 검출

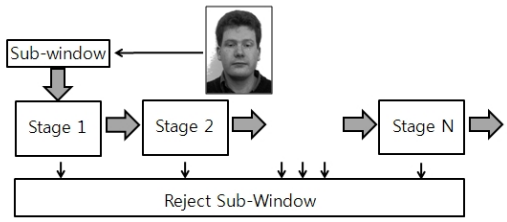
발화자 후보의 위치정보를 추정하기 위해선 3차원 공간상에 설치된 카메라의 기하학적 특성에 따라 좌, 우 카메라로부터 얻은 영상에서 상호간에 정합을 이루는 변이를 찾아내고 그 정보를 바탕으로 3차원 거리 정보를 추정하는 과정을 거친다. 우선 좌, 우 영상에서 기준이 될 특징으로서 사용할 얼굴을 검출하기 위해 Viola와 Jones가 제안한 Haar-like특징 기반의 Adaboost 알고리즘이 사용되었다[9]. Haar-like 특징을 이용한 얼굴 검출은 단순연산만을 사용한다. 다시 말해서 원래의 영상을 그대로 사용하지 않고 Haar-like 웨이블릿 특징을 이용하여 특정 영역안의 픽셀 합을 구하고 그 가중치를 곱한 합만을 계산한다.



(그림 2) Haar-like wavelet 특징

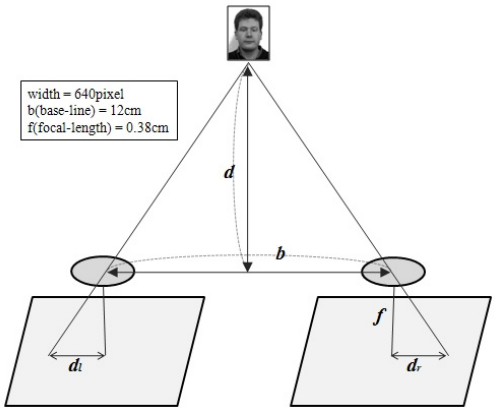
Adaboost 알고리즘은 얼굴검출을 위해 사용할 수 있는 주요한 특징들을 선택한 후 이런 특징들을 사용한 약한 분류기의 선형적인 결합을 통하여 강한 분류기를 생성하는 것이다[10]. 그림 2에 나타난 각각의 Harr-like 특징들이 약한 분류기가

되고 이들은 cascade구조를 형성함에 따라 강한 분류기가 된다. 입력영상에 따라 각 스테이지의 분류기로 얼굴과 얼굴이 아닌 부분의 분류를 순차적으로 수행하게 되며 각 단계의 분류기는 이전 스테이지를 통과한 학습데이터들을 사용한다. 그림 3은 단계적 얼굴분류기의 구조를 나타낸다.



(그림 3) 단계적 얼굴분류기의 구조

스테레오 카메라의 특징은 사람이 두 눈으로 물체를 입체시하고 원근을 판단하는 특징과 같다. 스테레오 영상에서 물체의 위치를 추적하는 방법에는 삼각 측량법을 사용한다.



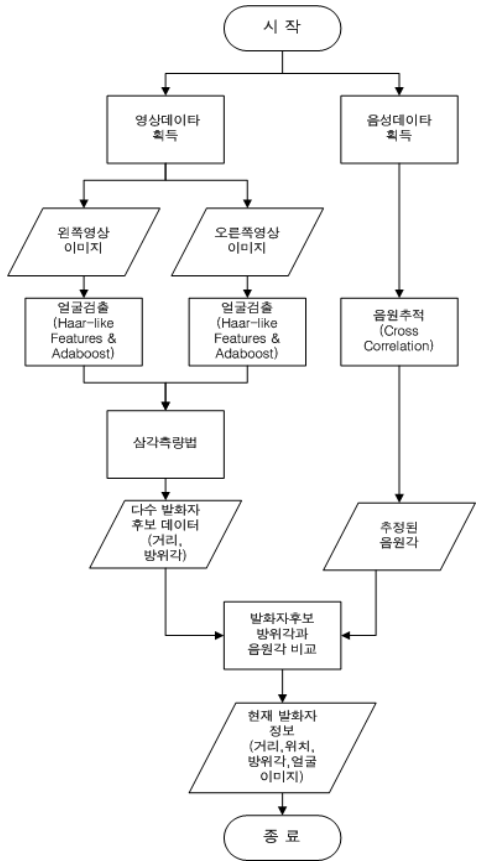
(그림 4) 스테레오 카메라의 구성도

삼각측량법을 이용한 거리계산은 식 7와 같다. 여기서 b는 좌, 우 카메라사이의 거리, d는 물체의 거리, f는 렌즈와 이미지 사이의 초점거리, $width$는 영상의 너비, 그리고 d_l과 d_r은 좌, 우 영상의 중심과 검출된 물체의 중심좌표사이의 거리이다.

$$d = \frac{b \times f}{d_l - d_r} \times width \tag{7}$$

3. 스테레오 시청각 기반의 화자 검출 시스템

3.1 스테레오 시청각 기반의 화자 검출 알고리즘



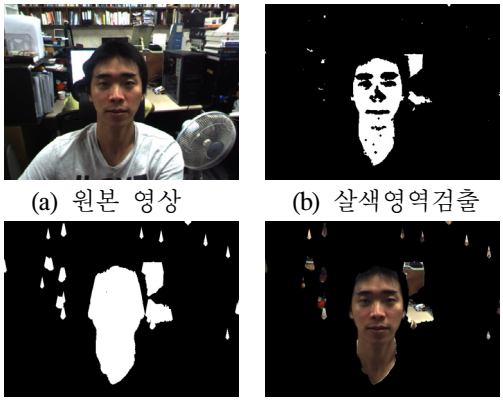
(그림 5) 스테레오 시청각 기반의 화자 검출 알고리즘

본 논문에서 제안한 스테레오 시청각기반의 화자 검출 알고리즘은 그림 5와 같다. 스테레오 카메라로부터 얻은 발화자 후보의 정보와 두 개의 마이크를 이용하여 음원 위치 추적 방법으로부터 얻은 정보를 비교하여 최종 발화자를 검출하게

된다. 우선 스테레오 카메라로부터 얻은 좌, 우 이미지에서 각각 Haar-like기반의 Adaboost알고리즘을 이용해 발화자 후보들의 얼굴을 검출하게 된다. 얼굴영역 검출의 성능 향상을 위하여 스테레오 카메라로부터 획득한 RGB 색상 모델을 YCbCr 색상 모델로 변환하고 살색 영역 검출과정을 거쳐 살색이 아닌 영역을 제거하였다. 피부색을 결정하는 임계치는 Cb와 Cr의 픽셀값의 분포를 분석하여 식 8과 같이 설정하였다.

$$SkinColor(x,y)=\begin{cases}255 & \text{if } (77 \le Cb \le 127) \cap (133 \le Cr \le 187) \\ 0 & Otherwise\end{cases} \quad (8)$$

살색 영역검출을 통해 이진화된 영상에서 얼굴 영역의 정확한 검출을 위해 침식(Erosion) 및 팽창(Dilation)연산을 이용하여 노이즈를 제거하고 원본영상과 AND연산을 수행하였고 그 결과는 그림 6과 같다.



(a) 원본 영상 (b) 살색영역검출

(c) 침식, 팽창연산 (d) AND연산

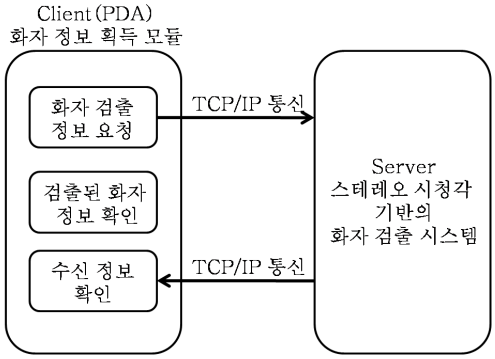
(그림 6) 피부색 영역 검출 전처리 과정

얼굴 영역 검출 후 삼각 측량법을 이용하여 각각 후보들의 거리와 클로즈업된 얼굴사진, 그리고 스테레오 카메라로부터의 방향각을 수집한다.

음원 위치 추정에서는 두 개의 마이크로부터 얻은 음성데이터로부터 CPSP기법을 이용하여 마이크의 중심으로부터 발화자가 발성한 음원의 방향각을 수집하였다. 최종적으로 영상데이터로부터 얻은 발화자 후보들의 방향각과 음성데이터로부터 얻은 음원의 방향각을 비교하여 다수 발화자 중 현재의 발화자 위치를 추정하게 된다.

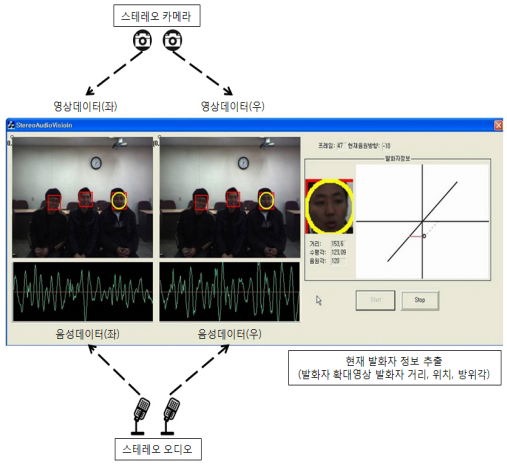
3.2 스테레오 시청각 기반의 화자 검출 시스템 구조

본 논문에서는 검출된 화자 정보를 바탕으로 보다 차별화된 서비스 제공을 위해 TCP 서버/클라이언트 구조 기반의 모바일 화자 검출 정보 획득 시스템을 구현하였고 구조는 그림 7과 같다. TCP기반의 서버/클라이언트 연결 지향형 소켓을 이용하여 모바일 기기에서 화자 검출 정보를 확인할 수 있도록 구성하였다.



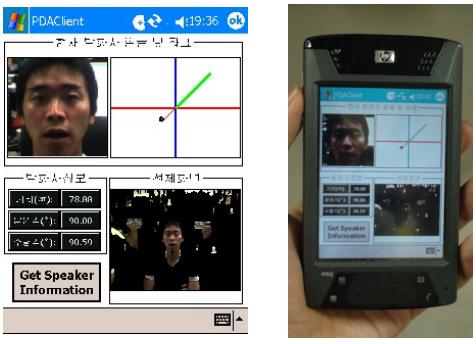
(그림 7) 모바일 화자 검출 정보 획득 시스템구조

서버 시스템의 구성은 개인용 컴퓨터(4GB RAM, Dual-Core CPU 2.8GHz)와 평행식 스테레오 카메라 Bumblebee2(Point Grey Inc.), 스테레오 오디오(Profire 2626, M-Audio)에 2개의 마이크로폰(DM-565, VASCOM)을 연결하여 구성하였다. 스테레오 카메라의 해상도는 640x480 pixel을 사용하였고 프레임율은 5.21frame/sec로 초기화 하였다.



(그림 8) 스테레오 시청각 기반의 화자 검출 서버 프로그램

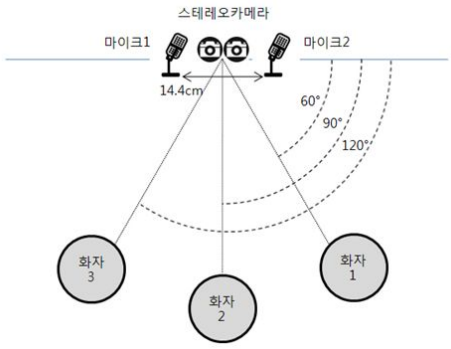
서버 시스템의 시뮬레이션 프로그램 구성은 그림 8과 같다. 스테레오 카메라로부터 얻은 좌, 우 영상이미지와 현재 좌, 우 마이크에서 측정되는 음성데이터를 확인할 수 있다. 그리고 제안한 화자 검출 알고리즘 과정을 거쳐 최종적으로 현재 발화자의 거리, 방향각, 클로즈업된 얼굴영상과 발화자의 위치를 화면에 나타내었다. 시뮬레이션 프로그램은 Visual Studio 2008을 이용하여 제작하였다.



(그림 9) 모바일 화자 검출 정보 획득 모듈

클라이언트 시스템의 구성은 휴대용 PDA(HP iPAQ hx4700(192MB 메모리, Intel PXA270 624MHz) OS는 Window Mobile 2003을 사용하였고 검출된 화자 검출 정보 획득 프로그램의 구성은 그림 9와 같다.

4. 실험 및 결과



(그림 10) 화자검출 실험환경

실험 환경은 그림 10과 같이 단순한 일반 강의실에서 20대 남성 3명을 대상으로 실험을 진행하였다. 본 실험에서 음성데이타의 sampling rate는 16kHz를 사용하였고 마이크간의 간격은 사람의 양쪽 귀간의 거리와 같은 14.4cm간격을 두었다. 이와 같은 조건에서 CPSP방법을 이용하여 지연시간을 구할 때 한 샘플이 지연될 경우 구분할 수 있는 방향각은 10° 이다. 그리고 스테레오 카메라의 focal length는 3.8mm, 시야각은 70° 이다. 그래서 피험자들이 모두 카메라 안에 위치할 수 있도록 피험자들을 카메라를 기준으로 각각 60°,90°,120° 에 위치하도록 하였고 번갈아가면서 10초가량 숫자를 세면서 발성을 하였다. 피험자들의 음성에 따라 인식률이 달라지는 현상을 고려하여 1회의 실험이 종료될 때마다 자리를 번갈아가면서 발성을 하였다. 실험은 측정거리를 달리하여 1m와 2m에서 진행하였으며 1회의 실험당 200프레임씩 총 10회에 걸쳐 각각의 실험마다 2000프레임씩 데이터를 수집하였다.

(표 1) 얼굴검출결과

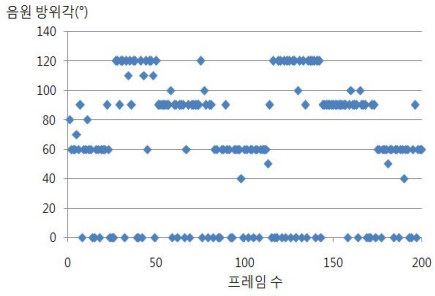
거리	얼굴 검출 결과			
	프레임 개수	성공	실패	검출율
1M	2000	1942	58	97.1%
2M	2000	1907	93	95.3%

표 1에서와 같이 단순한 강의실 배경에서의 발화자 후보들의 얼굴 검출율은 거리에 따라 다소 차이는 있지만 모두 95%이상으로 높게 나타나는 확인할 수 있다.

(표 2) 얼굴검출을 이용한 거리측정 결과

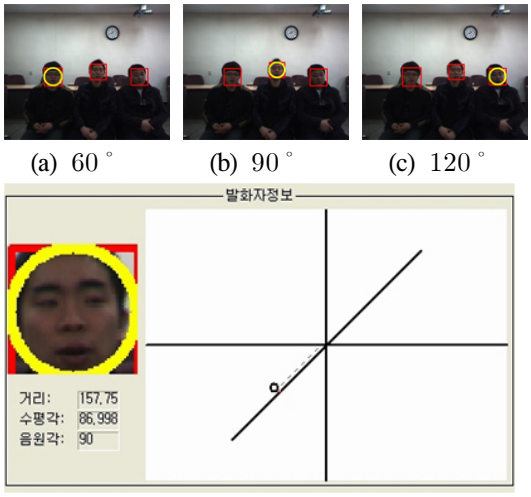
거리	거리 측정 결과		
	전체평균	실거리	표준편차
1M	101.5	100	2.8
2M	206.7	200	9.5

표 2는 얼굴 검출을 이용한 거리측정 실험으로 1m에서 표준편차가 2.8Cm 이고 2m에선 9.5Cm로 측정되어 거리가 멀어질수록 그 오차가 커지는 결과를 확인할 수 있었다.



(그림 11) 음원 위치 추정결과의 예

발화자를 검출하기 위한 데이터를 수집하기 위해 현재 발생한 음원이 잡음일 경우나 발생된 음원이 존재하지 않는 경우, 그리고 그 크기가 매우 미미한 경우를 제외하였다. 그리고 이러한 경우 그림 11에서와 같이 음원방위각이 0이라고 표시되고 최종 발화자 검출율 측정에서 제외하였다. 발화자 검출이 성공한 경우는 영상데이터에서 얻은 현재발화자의 방향각과 음원의 방향각을 비교하여 그 방향각의 오차가 $\pm 10^\circ$ 인 경우이며 실패한 경우는 현재 발화자가 아닌 후보자를 검출하거나 방향각의 오차가 $\pm 10^\circ$ 이상으로 발생하여 후보자를 검출하지 못하는 경우로 정의하였다.



(a) 60° (b) 90° (c) 120°

(d) 현재 발화자 정보(90°)

(그림 12) 스테레오 시청각 기반의 화자검출 결과

본 실험의 결과는 발화자 후보들의 방향각과 음원의 방향각을 비교하여 현재 발화자를 검출하게 된다. 발화자로 검출된 후보는 발화자 정보 창에 클로즈업된 얼굴 사진과 거리, 발화자의 위치가 나타나게 된다. 그림 12의 (a), (b), (c)와 같이 각기 다른 위치에 있는 피험자들은 적색 사각형으로 표시가 되고 발성을 할 때마다 실시간으로 현재의 발화자를 찾아 원으로 표시한다. 1m에서의 발화자 검출율은 86.3%이고 2m에서의 발화자 검출율은 80.57%로 거리가 멀어짐에 따라 검출율이 낮아지는 현상을 확인할 수 있었고 최종 발화자 검출율은 83.44%였다.

(표 3) 화자 검출 및 얼굴 검출 결과

실험 번호	발화자 검출 (1m)			
	프레임 개수	성공	실패	검출률
1	117	97	20	82.9%
2	131	113	18	86.2%
3	125	109	16	87.2%
4	140	123	17	87,8%
5	137	118	19	86.1%
6	138	115	23	83.3%
7	116	100	16	86.2%
8	142	126	16	88.7%
9	130	117	13	86.9%
10	136	120	16	88.2%
합 계	1312	1138	174	-
검출률	평균 발화자 검출율 (1m) : 86.3%			

실험 번호	발화자 검출 (2m)			
	프레임 개수	성공	실패	검출률
1	137	114	23	83.2%
2	131	106	25	80.9%
3	126	101	25	80.1%
4	135	110	25	81.4%
5	145	115	30	79.3%
6	135	107	28	79.2%
7	128	104	24	81.2%
8	138	109	29	78.9%
9	133	108	25	81.2%
10	140	112	28	80%
합 계	1312	1138	174	-
검출률	평균 발화자 검출율 (2m) : 80.57%			
최종 발화자 검출률 : 83.44%				

5. 결 론

본 논문에서는 스테레오 시청각 기반의 화자 검출 시스템을 제안하였다. 기존에 연구된 화자 검출 시스템들은 단일 발화자만을 대상으로 위치를 추정하거나[8], 다수 발화자를 대상으로 했을 경우엔 발화자의 얼굴과 방향각만을 검출할 수 있었다.[2] 하지만 본 시스템에서는 스테레오 카메라를 이용하여 발화자의 얼굴과 방향각뿐만 아니라 발화자의 정확한 위치까지 검출할 수 있었다. 실험 결과, 제안한 멀티모달 위치추정 시스템의 발화자 검출율은 83.44%로 기존에 연구된 시스템의 발화자 검출율인 65.5%보다 17.9%향상된 결과를 확인하였다[2]. 그리고 사용자에게 시간과 장소에 구애받지 않고 검출된 화자정보를 제공함으로서 능동적인 서비스가 가능할 수 있는 TCP기반의 서버/클라이언트 기반의 화자 검출 정보 획득 서비스를 제공할 수 있도록 하였다. 그 결과 모바일 환경에서 사용하기 힘든 스테레오 시청각 시스템을 분산처리를 활용하여 실행 속도 및 하드웨어 제약사항을 극복하였다. 이렇게 향상된 위치정보의 신뢰성을 바탕으로 추출된 발화자의 얼굴 영상데이터 이용하여 발화자 구분이 이루어진다면 향후 사용자의 위치정보를 이용한 적합한 서비스를 제공할 수 있을 것이다. 향후 영상감시 및 방범 시스템이나 컨퍼런스나 각종 회의에서 스테레오 원격제어 시스템 등에 응용될 수 있을 것이다.

참고문헌

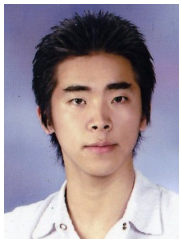
[1] A. Kushal, M. Rahurkar, Li Fei-Fei, J. Ponce, T. Huang, “Audio-Visual Speaker Localization Using Graphical Models” 18th International Conference on Pattern Recognition. Vol 1, 2006 pp.291-294

[2] T. Takiguchi, J. Adachi, Y. Ariki, “Audio-Based Video Editing with Two-Channel Microphone” International Conference on Multimedia and Ubiquitous Engineering. 2008. pp.282-287

[3] H.Atmoko, D.C.Tan, G.Y.Tian, Bruno Fazenda, “Accurate Sound Source Localization in a Reverberant Environment using Multiple Acoustic Sensors”, Measurement Science and

Technology Journal, Vol.19 No.2, 2008
[4] K. Nakadai, H. G. Okuno, H. Kitano, “Real-time Sound Source Localization and Separation For Robot Audition″ IEEE International Conference on Spoken Language Process. 2002. pp.193-196
[5] M. Omologo, P. Svaizer, “The generalized correlation method for estimation of time delay″, IEEE Transactions. Acoustics. Speech and signal Processing, Vol 25, No 4, 1976
[6] B.C. Park, K.D. Ban, K.C. Kwak, H.S. Yoon, “Sound Source Localization Based on Audio-visual Information for Intelligent service Robot″, The 8th International Symposium on Advanced Intelligent Systems. 2007. pp.515-519
[7] 진상현, 김동주, 홍광석, “스테레오 비전 기반의 사용자 위치정보 추정 방법에 관한 연구” 한국 신호처리 시스템학회 추계 학술대회 논문집. 제9권 제2호 pp.353-356
[8] 박정옥, 나승유, 김진영, “휴모노이드 로봇을 위한 시청각 정보 기반 음원 정위 시스템 구현” 한국음성학회, 음성과학 제11권 4호, 2004. pp.29-42
[9] Paul Viola, Michael Jones. “Rapid Object Detection using a Boosted Cascade of Simple Features″, IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, 2001. pp.511-518
[10] 채영남, 정지년, 양현승. “얼굴 색상과 에이다부스트를 이용한 효율적인 얼굴 검출”, 정보과학회논문지 소프트웨어 및 응용 제36권 제7호, 2009. pp 548-559

◐ 저 자 소 개 ◑



안 준 호 (Jun-ho An)
2010년 광운대학교 전파공학과 졸업(학사)
2010년~현재 성균관대학교 휴대폰학과 석사과정
관심분야 : HCI, 패턴인식, 모바일, 영상처리
E-mail : amadasv@skku.edu

홍 광 석 (Kwang-Seok Hong)
1985년 성균관대학교 전자공학과 졸업(학사)
1988년 성균관대학교 대학원 전자공학과 졸업(석사)
1992년 성균관대학교 대학원 전자공학과 졸업(박사)
1990~1993년 서울보건대학 전산정보처리과 전임강사
1993~1995년 제주대학교 정보공학과 전임강사
1995년~현재 성균관대학교 정보통신공학부 교수
관심분야 : 오감인식, 융합 및 재현, HCI
E-mail : kshong@skku.ac.kr