

3D 거리 센서를 이용한 강의용 광역 마이크 시스템

Widerange Microphone System Using 3D Range Sensor

Woojin Oh*

*Professor, School of Electronic Engineering, Kumoh National University of Technology, Gumi, 39177 Korea

ABSTRACT

In this paper, 3D range sensor is applied to the sensor-based widerange microphone system for lectures. Since the 2D range sensor measures the shortest distance of the speaker, an error occurs and the performance is degraded. The 3D sensor provides a 160x60 distance image so that the position of the speaker can be obtained with accuracy. We propose a method for obtaining the distance per pixel required to determine the absolute position of the speaker from the distance image. The proposed array microphone system using the 3D sensor shows the improvement of 0.8~1.5dB compared to the previous works using 2D sensor.

Keywords : Wide range microphone, 3D range sensor, Array microphone, Range image

I. 서 론

일반적인 광역 마이크 시스템은 넓은 마이크 수신 영역을 위해 다수의 배열 마이크에 복잡한 신호처리 알고리즘을 적용해야 한다. 강의용 광역 마이크는 수신 영역이 지정된 영역에 소수의 화자만 존재하는 특수한 환경이므로 구현이 간단해질 수 있다 [1-3].

논문 [2]에서 거리 센서로 FMCW(Frequency Modulated

Continuous Waveform) 레이더를 사용하여 화자의 위치를 추적하고 4개의 배열 마이크 신호를 지연-합(delay-sum)으로 처리할 수 있음을 보였다. 이 방식은 기존의 배열 마이크 기반의 위치 추적 방식인 논문 [3]과 유사한 성능을 보이면서도 배열 신호 처리에 필요한 화자 검출과 추적 알고리즘이 제거되어 구현이 간단한 장점이 있다. 소수의 화자가 있는 경우에도 각각의 위치를 검출하여 신호의 크기에 따라 적용하는 것도 가능할 것이다.

본 논문에서는 강의용 광역 마이크 시스템에 여러 가지 거리 센서의 특성을 비교하고 적합한 센서를 제안하고자 한다. 2장에서 2D 거리 센서를 이용한 강의용 배열 마이크 시스템에 대하여 알아보고 3장에서 3D 센서를 분석하고 이를 적용한 성능을 제시하고 4장에서 결론을 맺는다.

II. 2D 거리 센서 기반의 강의용 광역 마이크

일반적인 강의용 마이크 시스템은 탁상형 유선 마이크나 휴대용 또는 부착형 무선마이크를 많이 사용한다. 광역 마이크 시스템은 천정에 배열 마이크를 설치하여 마이크 관리가 편리하면서 이동성을 제공하는 방식이다. 원거리 수신성능은 화자 추적과 배열 마이크의 신호 처리로 얻게 된다 [1-2].

화자 추적은 성능에 영향을 미치는 중요한 요소로 배열 마이크 수신된 각각의 신호에서 구해야 한다. 각 마이크에 수신된 음성 신호의 시간 지연, 전력 크기 등을 이용하여 위치를 찾는 여러 방안이 연구되어왔다. 화자 추적에는 무선통신처럼 랜덤한 신호를 사용할 수 없으므로 음성 신호에서 구해야 한다. 그러나 음성 신호는 낮은 주파수 대역에서 존재하여 주파수 변화가 크고, 공명 주파수를 갖는 포먼트(formant) 특성으로 인하여 상관(correlation) 방식으로 실시간 위치 추적이 복잡해진다. 배열 마이크의 간격은 신호의 반 파장이 되어야 하지만 음성의 파장은 수 cm에서 수 m의 범위이므로 단일

Received 19 August 2021, Revised 23 August 2021, Accepted 28 August 2021

* Corresponding Author Woojin Oh(E-mail:wjoh@kumoh.ac.kr, Tel:+82-54-478-7485)

Professor, School of Electronic Engineering, Kumoh National Institute of Technology, Gumi, 39177 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.10.1448>

print ISSN: 2234-4772 online ISSN: 2288-4165

값으로 설정하기 어렵다. 강의용 광역 마이크는 수신 범위가 제한되므로 화자의 위치를 음성 신호로 추적하는 것보다 거리 센서를 사용하는 것이 가능하며 정확도도 높다.

논문 [2]는 배열 마이크용 거리 센서로 FMCW 레이더를 제안하고 실험으로 성능 개선을 확인한 바 있다. 이 레이더 센서는 장애물에 대한 거리와 방위각을 제공하는 2D 센서이다. 화자를 2D 평면에서 측정하기 때문에 실제 화자의 위치가 다를 수 있다는 것이다. 레이더 센서의 고도각(elevation angle)이 $\pm 14^\circ$ 이므로 2m 거리에서 동일 방위각에 존재하는 장애물이 그림 1과 같이 2m에서 2.26m의 거리로 측정될 수 있어 최대 26cm의 오차를 가질 수 있는 것이다. 이 문제를 해결하기 위해 공간 위치를 확인할 수 있는 3D 센서를 적용하였다.

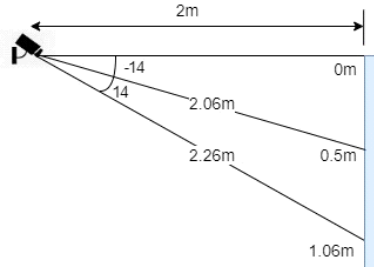


Fig. 1 Distance variation according to the range of elevation angle

III. 3D 거리 센서 기반의 분석 및 실험

3D 거리 센서의 동작 방식은 크게 2가지로 구분되며 사람의 눈처럼 스테레오 카메라로 시차를 이용한 수동형 측정방식과 레이저나 적외선을 투사하는 능동형 측정방식이 있다. 수동형 방식은 수십 m급의 원거리 탐색에 사용되며 크거나 정밀도의 한계로 정밀 측정에는 잘 사용하지 않는다. 능동 방식은 레이저나 적외선을 투사하여 반사된 빛의 시간지연, 위상차, 구조광(structured light) 등을 이용하여 거리를 구하는 것이다. 이 방식은 소형화, 저전력 등이 가능하여 최근 스마트폰의 얼굴인식 및 동작인식, 자동차용 전방 센서 등에 많이 사용되고 있다 [4-5].

광역 마이크에 적합한 3D 센서는 3m 이내의 근거리를 지원하고 소형 및 저가이어야 한다. 고성능 원거리를 지원하는 레이저 기반의 LiDAR는 고가이므로 제외한다.

광역 마이크용 센서는 능동형으로 적외선을 사용하는 TOF(Time-Of-Flight)방식이나 구조광 방식이 될 것이다. 구조광 방식은 3D 스캐너, 콘솔 게임기 등에 적용되고 있으며 실내 환경에 유용하지만 투사 광원의 세기와 외부 환경의 밝기에 따라 정밀도가 영향을 받으므로 부적합하다.

본 논문에서는 3D 센서로 위상차 기반의 간접 TOF를 이용하는 능동형 적외선 센서인 CygLiDAR D1을 사용하였다. 국내에서 개발된 2D/3D 복합 센서로 소형, 저가이지만 808nm 파장의 적외선 LED와 레이저 다이오드를 내장하여 근거리용 3D와 원거리용 2D 측정을 동시에 제공한다. 센서의 주요 제원은 다음과 같다 [6].

- 측정 거리: 20~800cm (2D), 5~200cm (3D)
- 오차 및 분해능: $\pm 1\%$, 160x60 (3D)
- 시야각: 수평 120° , 수직 65°
- 크기 및 무게: 37.4x37.4x24.5mm, 28g

그림 2는 화자와 수평으로 2m, 수직으로 2.6m에 3D 센서를 설치했을 때의 거리 영상(depth image)이다. 그림 2(a)와 같이 상단의 물체는 천장이고, 하단 중앙부는 교탁으로 항상 측정되는 고정 물체이며, 중간 부분은 화자와 칠판이 검출된 것이다. 이는 최대 측정 거리를 얻기 위해 센서의 문턱값을 낮추었기 때문이다. 강연과 같은 고정 환경에서 사람과 같은 큰 장애물을 검출하는 경우에 3.2m까지 검출할 수 있음을 확인하였다. 그림 2(b)는 화자가 왼쪽에 있는 경우이며 좌표 (45, 37)의 거리는 2.502m이다. 그림에서 검출된 모양이 다른 것은 화자와 적외선 LED의 각도가 수직에 가까울수록 많이 반사되기 때문이다.

거리 영상의 좌표로부터 절대 위치를 알기 위해서는 화소당 거리를 알아야 한다. 첫 번째 방법은 거리 영상의 중앙점 (60,30)을 이용하는 것이다. 그림 3(a)는 장애물이 중앙점에 있을 때의 도식이다. 센서와 중앙의 장애물이 수평으로 1.75m, 거리가 2.16m로 알고 있다면 다른 중앙점은 측정된 거리 영상로부터 화소당 거리를 구할 수 있다. 중심선 각도 θ 는 $\cos^{-1}(1.75/2.16)=35.9^\circ$ 이므로 임의의 중심점에서 거리가 2.5m로 측정되면 수직 거리 $d=1.47m$ 를 얻을 수 있다. 이때의 화소당 거리는 $1.47/30=0.049m/pixel$ 이 된다. 이 방법은 강의 환경에서 화자가 중앙에서 시작하는 경우가 많으므로 타당한

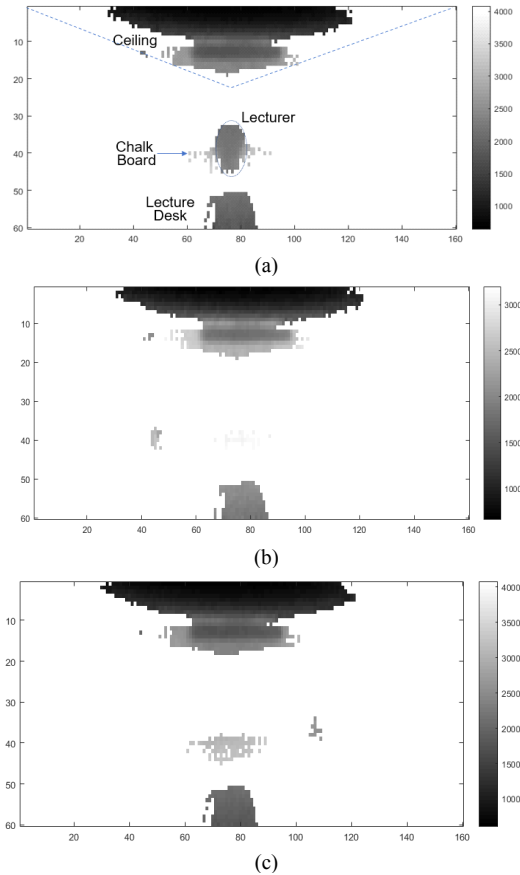


Fig. 2 Range images of 3D sensor for Lecturer at (a) center, (b) left, and (c) right

방법으로 보인다.

두번째 방법은 그림 3(b)처럼 동일 축 상에 있는 두 점을 이용하는 것이다. 화자가 동일 y축에서 A, B 두 지점을 이동한다고 가정하면 삼각함수 공식에 의해 다음의 관계를 갖는다.

$$x^2 + (aP)^2 = r_1^2 \quad (1)$$

$$x^2 + (bP)^2 = r_2^2 \quad (2)$$

여기서 P 는 화소당 거리이고 수식을 정리하면 다음과 같다.

$$P = \sqrt{\frac{r_1^2 - r_2^2}{a^2 - b^2}} \quad (3)$$

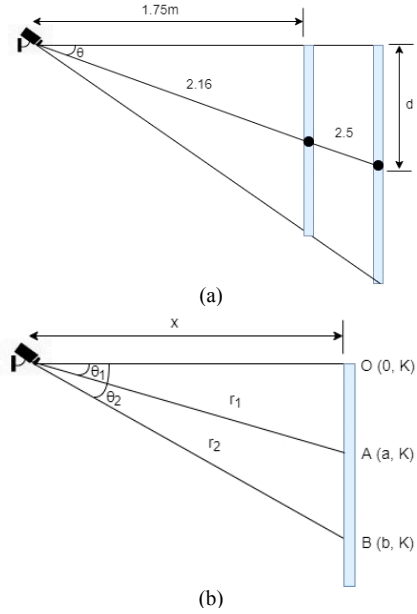


Fig. 3 Calculation of Distance per pixel: (a) Center point method, (b) Two points method

화자가 교탁에서 좌우로 이동하면 y축이 고정되고 앞뒤로 이동하면 x축이 고정된다. 화자의 위치에 따라 P 가 변화하므로 두 방식을 적용하여 P 를 구하여 절대 위치를 알 수 있다.

실험은 마이크를 교단에서 수평으로 1.3m 떨어진 위치에 50cm 간격으로 4개를, 센서는 화자가 확보되도록 수평으로 2.1m 위치에 설치하고 진행하였다. 센서와 마이크의 높이는 2.6m이며 배열 마이크의 음성 결합은 다음과 같이 지연-합을 사용하였다. [7]

$$y(t) = \sum_{i=0}^3 s_i m_i(t - d_i(t)) \quad (4)$$

여기서 $m_i(t)$ 는 i 번째 마이크 신호이고 s_i 는 마이크 신호의 반전을 보상하는 부호 값이며 $d_i(t)$ 는 거리 차에 의한 지연이다.

표 1은 교단의 영역인 좌우로 2m씩 화자가 이동하며 측정한 결과이다. 배열 신호처리 없이 가장 큰 전력의 마이크 신호로 처리한 경우, 이전의 2D 센서를 사용한 경우, 제안한 3D 센서 방식의 성능을 비교하였다. 최대치 방식보다 평균 9.8dB, 2D 센서 방식보다 0.8~1.5dB 개선되었다. 중앙부의 성능 개선이 상대적으로 낮은 것

은 배열 처리가 없을 때의 값이 크기 때문으로 보인다.

Table. 1 Performance comparison in various position

Method \ Position	-2 m	0 m	2 m
None: $\max(P_{m_i})$	1.81E-4	5.06E-4	2.28E-4
2D Sensor [2]	0.0019 (10.2dB)	0.0026 (7.1dB)	0.0016 (8.5dB)
Proposed 3D Sensor	0.0025 (11.4dB)	0.0031 (7.9dB)	0.0023 (10.0dB)

제안된 센서는 이전의 2D FMCW 레이더보다 저렴하면서도 우수한 성능을 확인하였다. 이 센서는 2D와 3D가 복합되어 근거리와 원거리에서 동시에 활용이 가능한 장점이 있다.

IV. 결 론

본 논문에서는 센서 기반의 강의용 광역 마이크 시스템에 3D 위치 센서를 적용하였다. 2D 센서는 화자의 최단 거리가 측정되어 실제 위치와 차이가 발생하여 성능이 저하된다. 적용한 3D 센서는 160x60의 거리 영상으로 측정하므로 화자의 정확한 발성 위치를 얻을 수 있다. 거리 영상에서 장애물의 위치로부터 화소당 거리를 구하는 방법을 제시하고 배열 마이크에 지연 보상하여 성능 개선을 확인하였다. 제안된 3D 센서 방식은 2D 센서보다 0.8~1.5dB의 성능이 개선되는 것을 확인하였다.

ACKNOWLEDGEMENT

This paper was supported by sabbatical year program of Kumoh National University of Technology.

REFERENCES

- [1] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*, New York, NY: Springer, 2008.
- [2] W. Oh, "Widerange microphone system for lecture using FMCW radar sensor," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 4, pp. 611-614, Apr. 2021.
- [3] W. Oh, "Wide Coverage Microphone System for Lecture Using Ceiling-Mounted Array Structure," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 4, pp. 624-633, Apr. 2018.
- [4] M. Hansard, S. Lee, S. O. Choi, and R. Horaud, *Time-of-flight cameras: principles, methods and applications*, New York, NY: Springer Science & Business Media, 2012.
- [5] B. Park, D. Kim, and Y. Seo, "A method of improving the quality of 3D images acquired from RGB-depth camera," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 5, pp. 637-644, May. 2021.
- [6] 2D/3D Dual CygLiDAR D1 User Manual, Ver 1.0.9, [Internet] Available: <https://www.cygbot.com/downloads>.
- [7] S. Seo and M. Kim, "3D Impulsive Sound-Source Localization Method through a 2D MEMS Microphone Array using Delay-and-Sum Beamforming," in *Proceedings of the 9th International Conference on Signal Processing Systems (ICSPS 2017)*, Auckland, New Zealand, pp. 170-174, 2017.