

Steered Response Power를 위한 검색 공간 클러스터링 연구

정재연, 육동석*
고려대학교 컴퓨터학과

A Study of Search Space Clustering Algorithm for Steered Response Power

Jaeyoun Chung, Dongsuk Yook

Department of Computer Science and Engineering, Korea University

E-mail : tochan@voice.korea.ac.kr, yook@voice.korea.ac.kr

Abstract

Steered response power(SRP) based algorithm uses a focused beamformer which steers the array to various locations and searches for a peak in output power to localize sound sources. SRP-PHAT, a phase transformed SRP, shows high accuracy, but requires a large amount of computation time. This paper proposes an algorithm that clusters search spaces in advance to reduce computation time of SRP based algorithms.

I. 서론

최근 여러 개의 마이크로폰을 이용하여 음향의 위치를 추적하는 연구가 많이 이루어지고 있다. 이러한 음향 위치 추적 기술은 화자의 위치 정보를 이용하여 화자에게 자동으로 초점을 맞추는 화상 회의 시스템[1], 화자의 위치 정보를 이용하여 강화된 음성을 사용하는 음성 인식 시스템[2], 그리고 사용자의 위치 정보를 이용하여 사용자에게 편리한 서비스를 제공하는 휴머노이드 로봇[3] 등 다양한 분야에 적용될 수 있다.

음향 위치 추적 기술은 일반적으로 세 가지 종류로 분류할 수 있다. 첫 번째는 조향된 빔 형성기(steered

beamformer)를 이용한 방법이다. 대표적인 예로 steered response power(SRP) 방법이 있다[4]. 두 번째는 고해상도 스펙트럼 추정(high resolution spectral estimation)을 이용한 방법이다. 대표적인 예로 multiple signal classification (MUSIC) 방법이 있다[5]. 그리고 세 번째는 도착 지연 시간(time delay of arrival)을 이용한 방법이다. 대표적인 예로 generalized cross correlation(GCC) 방법이 있다[6]. 그 중 SRP 기반 방법이 MUSIC 기반 방법과 GCC 기반 방법에 비해 정확도가 높지만[4][7], 음향의 위치를 추적하기 위해 모든 구간에 대해 빔을 형성하여 빔 형성기의 출력 파워가 가장 높은 구간을 검색하기 때문에 계산량이 많다. 이를 해결하기 위한 연구들이 이뤄지고 있지만 [4][8] 주로 음향이 원거리에 있음을 가정하여 위치가 아닌 방향을 찾는 원거리 SRP 기반 방법에서의 연구가 이뤄지고, 위치를 찾는 방법에서 검색 시간을 줄이는 연구는 미비하다. 본 논문에서는 3차원 공간에서 위치를 찾는 SRP 기반 방법에서 계산량을 줄이기 위해 기존의 블록 단위로 모든 공간을 검색하는 방법을 개선한 빔 형성기의 출력 파워가 같은 블록 즉, 마이크로폰 간의 지연 시간 조합이 같은 블록끼리 클러스터링 하는 검색 공간 클러스터링 방법을 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 신호 모델링, SRP 기반의 음향 위치 추적 방법, 그리고 검색 공간 클러스터링 방법을 다루고 3장에서는 2장에서의 검색 공간 클러스터링 방법을 이용하여 실험한 결과들을 비교하고, 4장에서 결론을 맺는다.

* 교신저자

본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업의 연구결과로 수행되었습니다.

II. SRP 기반의 음향 위치 추적 및 검색 공간 클러스터링 방법

2.1 신호 모델링과 교차상관

$x_n(t)$ 를 n 번째 마이크로폰에서 받은 신호라고 하면 반향, 잡음, 그리고 지연 시간을 고려한 신호 모델링은 다음과 같다.

$$x_n(t) = h_n(t) * s(t - \tau_n) + n_n(t) \quad (1)$$

τ_n 은 음향의 위치와 마이크로폰 간의 거리에 의한 지연 시간, $h_n(t)$ 는 반향에 해당하는 충격 응답(impulse response), $s(t)$ 은 원본 음향 신호, $n_n(t)$ 는 비상관 잡음(uncorrelated noise), 그리고 $*$ 는 선형 승적(linear convolution)을 나타낸다. n 번째 마이크로폰과 m 번째 마이크로폰에서 받아들인 신호 간의 파워스펙트럼(power spectrum) $G_{x_n x_m}(\omega)$ 를 이용한 교차상관(cross correlation) $R_{x_n x_m}(\tau)$ 은 다음과 같다.

$$G_{x_n x_m}(\omega) = H_n(\omega) H_m^*(\omega) G_{s s}(\omega) e^{j\omega(\tau_n - \tau_m)} \quad (2)$$

$$R_{x_n x_m}(\tau) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(\omega) G_{x_n x_m}(\omega) e^{j\omega\tau} d\omega \quad (3)$$

$H_n(\omega)$ 과 $H_m(\omega)$ 는 각각 $h_n(t)$ 과 $h_m(t)$ 의 푸리에 변환(the Fourier transform), $*$ 는 켈레복소수(complex conjugate)변환, 그리고 $\psi(\omega)$ 는 $H_1(\omega)H_2^*(\omega)$ 의 역수로서 주파수 가중함수를 나타낸다.

2.2 SRP 기반의 음향 위치 추적 방법

N 개의 마이크로폰을 이용하여 소리의 신호를 받아들였을 때 빔 형성기가 초점을 맞춘 위치 \vec{q}_s 와 n 번째 마이크로폰 간의 거리에 의한 지연 시간 τ_n 만큼 마이크로폰의 신호를 이동하고, 채널 잡음에 강인하도록 필터를 적용하여 주파수 영역에서 N 개의 신호 모두를 더한 filter-sum 빔 형성기의 출력 $Y(\omega, \vec{q}_s)$ 과 출력 파워 $P(\vec{q}_s)$ 는 다음과 같다.

$$Y(\omega, \vec{q}_s) = \sum_{n=1}^N H_n(\omega) X_n(\omega) e^{j\omega\tau_n} \quad (4)$$

$$P(\vec{q}_s) = \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega \quad (5)$$

$H_n(\omega)$ 과 $X_n(\omega)$ 는 각각 필터와 마이크로폰에서 받은 신호의 푸리에 변환을 나타낸다. 식 (3)과 식 (4)를 이용하여 빔 형성기의 출력 파워인 식 (5)를 교차상관으로 나타내면 다음과 같다.

$$P(\vec{q}_s) = 2\pi \sum_{n=1}^N \sum_{m=1}^N R_{x_n x_m}(\tau_{mn}) \quad (\tau_{mn} = \tau_m - \tau_n) \quad (6)$$

N 개의 마이크로폰 간의 지연시간 차의 조합 ($\tau_{12}, \tau_{13}, \dots, \tau_{M(N-1)}$)은 $N(N-1)/2$ 개다. 지연시간 차 τ_{mn} 은 다음의 식으로 구할 수 있다.

$$\tau_{mn} = \frac{|\vec{m}_m - \vec{q}_s| - |\vec{m}_n - \vec{q}_s|}{c} \quad (7)$$

c 는 소리의 속도, \vec{m}_m 과 \vec{m}_n 는 각각 m 번째 마이크로폰과 n 번째 마이크로폰의 3차원 공간에서의 위치를 나타낸다. 식 (6)의 빔 형성기의 출력 파워를 이용한 SRP 기반 방법의 위치 추적 식은 다음과 같다.

$$\hat{\vec{q}}_s = \underset{\vec{q}_s}{\operatorname{argmax}} P(\vec{q}_s) \quad (8)$$

앞의 식들을 이용하여 SRP 기반 방법으로 3차원 공간에서 음향의 위치 $\hat{\vec{q}}_s$ 는 다음의 과정으로 구할 수 있다.

표 1. SRP 기반 방법의 3차원 음향 위치 추적 과정

- 1) 식 (3)을 이용하여 마이크로폰에서 받아들인 신호 간의 교차상관을 구한다.
- 2) 3차원 공간에서 음향이 위치할 수 있는 모든 후보 위치 \vec{q}_s 에서 과정 3)을 실행한다.
- 3) 식 (6)을 이용하여 위치 \vec{q}_s 로 빔을 형성하고, 위치 \vec{q}_s 와 마이크로폰 간의 지연시간 차 τ_{mn} 을 교차상관 $R_{x_n x_m}(\tau)$ 에 대입하여 더함으로 빔 형성기의 출력 파워 $P(\vec{q}_s)$ 를 구한다.
- 4) 식 (8)을 이용하여 3차원 공간의 모든 후보 위치 중 빔 형성기의 출력 파워가 가장 큰 위치 $\hat{\vec{q}}_s$ 를 구한다.

2.3 검색 공간 클러스터링 방법

3차원 공간에서 음향의 위치를 찾는 SRP 기반 방법에서 기존의 음향의 후보 위치 검색 방법[4]은 정해진

일정한 간격([4]에서는 10cm)으로 후보 위치 블록을 나누고, 각각의 후보 위치 블록에서 후보 위치와 마이크로폰 간의 지연시간 차를 미리 구한 뒤 이를 이용하여 빔 형성기의 출력 파워를 구하였다. 그렇지만 후보 위치 블록 중 지연시간 차의 조합이 같은 경우, 빔 형성기의 출력 파워가 같기 때문에 이러한 중복 연산을 없앴으므로 SRP 기반 방법의 계산량을 감소할 수 있다. 표 2의 검색 공간 클러스터링 방법을 이용하여 후보 위치 중 지연시간 차의 조합이 같은 경우의 중복 연산을 없애 계산량을 줄인다.

표 2. 검색 공간 클러스터링 과정

- 1) 음향이 위치할 수 있는 모든 3차원 공간에서 정해진 간격으로 후보의 위치 블록을 나눈다.
- 2) 각 블록의 무게 중심을 후보 위치 좌표로 설정한다.
- 3) 식 (7)을 이용하여 모든 후보 위치 블록에서 $N(N-1)/2$ 개의 지연시간 차 조합을 구한다.
- 4) 지연시간 차 조합이 같은 후보 위치 블록들을 묶어 하나의 새로운 후보 위치 블록을 만든다.
- 5) 묶여진 위치 블록의 무게 중심을 새로운 후보 위치 블록의 좌표로 설정한다.

III. 실험 결과

검색 공간 클러스터링 방법을 적용한 SRP 기반 방법을 이용하여 휴머노이드 로봇 환경에서 3차원 공간 음향 위치 추적 실험을 하였다.

방의 크기는 7m×5m×3m로 설정하였고, 휴머노이드 로봇의 몸통을 원통으로 가정하여 지름 0.3m, 높이 0.45m인 원통의 윗면과 아랫면에 원 둘레를 따라 각각 8개의 마이크로폰(총 16개의 마이크로폰)을 등 간격으로 배치하였다. 7m벽을 x축, 5m벽을 y축, 3m벽을 z축으로 하여 좌측 하단 바닥을 원점으로 아랫면 원통의 중심이 (1m, 1m, 0.95m), 윗면 원통의 중심이 (1m, 1m, 1.4m)가 되도록 원통을 놓았다. 소리가 난 위치에 따라 소리가 직접 도착하는 마이크로폰과 원통을 둘러 도착하는 마이크로폰으로 나뉘지기 때문에 음향의 후보 위치에 따라 지연시간 차 조합에 사용되는 마이크로폰들의 조합이 바뀌는 것을 고려하여 실험에 사용한 마이크로폰과 후보 위치 검색 공간을 다음과 같이 설정하였다. 그림 1은 원통의 윗면을 위에서 바라본 그림으로 8개의 마이크로폰 배치와 직접 소리가 도착하는 마이크로폰의 구성에 따라 16 구간으로 나눈 음향의 후보 위치 구간을 나타낸 그림이다. 원 둘레의 점

들이 각각의 마이크로폰이고, M1~M8의 번호를 붙였다. 구간 1에서 난 소리는 M1, M2, M3의 마이크로폰에 직접 들어오고, 구간 2에서 난 소리는 M1, M2, M3, M4의 마이크로폰에 직접 들어온다. 이와 같이 홀수 구간은 3개의 마이크로폰에, 짝수 구간은 4개의 마이크로폰에 소리가 직접 들어온다. 이를 원통의 아랫면까지 확장하면 홀수 구간과 짝수 구간에서 난 소리는 각각 6개와 8개의 마이크로폰에 직접 들어온다.

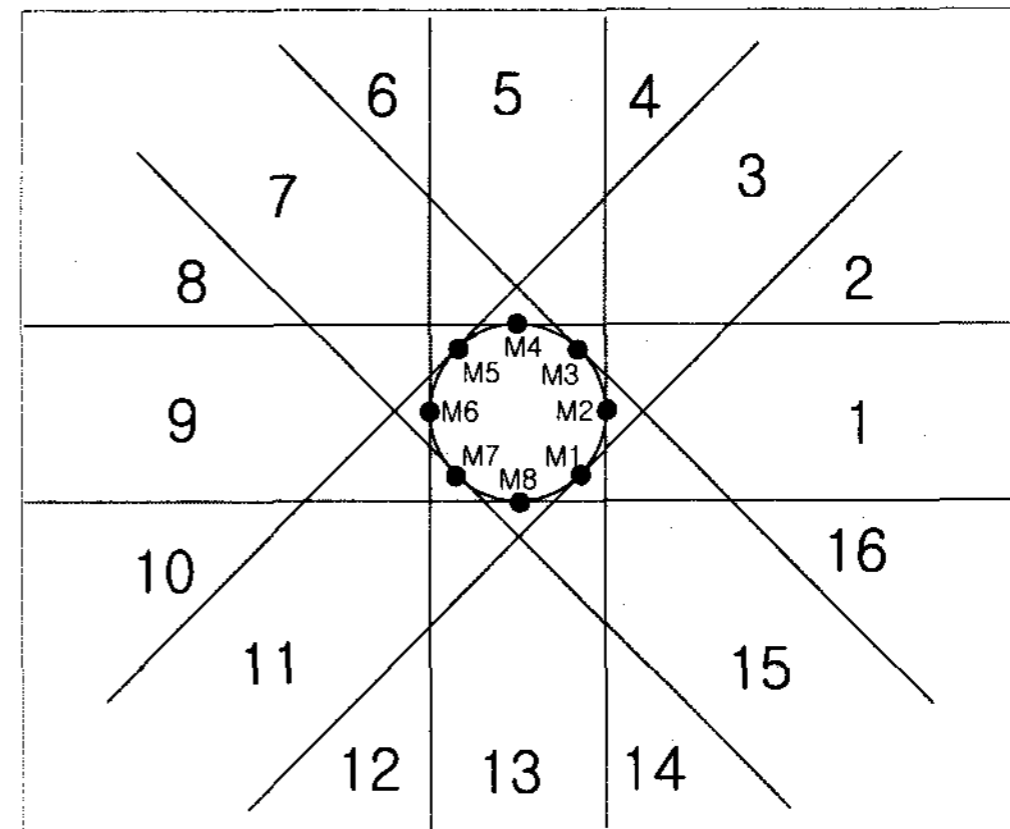


그림 1. 마이크로폰 배치와 후보 위치 구간

또한, 구간 1의 후보 위치 블록의 개수는 홀수 구간에서의 것과 구간 2의 후보 위치 블록의 개수는 짝수 구간에서의 것과 같기 때문에, 구간 1과 구간 2에서의 후보 위치 블록 개수의 합을 알면 전체 구간의 개수를 알 수 있다. 따라서 본 논문에서는 원통의 윗면에서는 M1, M2, M3, M4 마이크로폰, 아랫면에서는 같은 x, y 좌표를 갖는 마이크로폰, 총 8개의 마이크로폰을 이용하였고, 3차원 위치 추적을 위한 후보 위치는 x축과 y축은 구간 1과 구간 2로, z축은 0.5m부터 2.5m의 구간으로 설정하여 실험하였다. 샘플링 주파수는 16KHz로, 윈도우의 크기는 512 샘플(32 ms)로 두고, 연산 시간은 Intel(R) Pentium(R) M processor 1.86GHz의 cpu와 512MB의 RAM 사양의 PC에서 SRP 기반 방법을 이용하여 20개 윈도우에서 3차원 공간에서 음향 위치 추적을 실행한 평균 연산 시간을 측정하였다. 블록의 크기는 20cm, 15cm, 10cm, 5cm, 3.3cm로 설정하였다.

표 3은 블록 크기에 따른 구간 1, 2에서의 검색 공간 클러스터링 적용 전후의 후보 위치 블록의 개수를 비교한 표이다. 검색 공간 클러스터링을 적용함으로써 후보 위치 블록의 개수가 감소하였다. 특히 블록의 크기가 작아질수록 클러스터링 적용에 따른 후보 위치 블록 개수의 감소율이 커졌다.

표 4는 블록 크기에 따른 구간 1, 2에서의 검색 공간 클러스터링 적용 전후의 연산 시간을 비교한 표이

다. 검색 공간 클러스터링을 적용함으로써 연산 시간이 감소하였다. 또한 블록의 크기가 작아질수록 클러스터링 적용에 따른 연산 시간 감소율이 커졌다. 3.3cm의 클러스터링 전 연산 시간이 큰 이유는 메모리 사용량 증가로 인해 PC의 전체적인 속도가 저하된 것이 원인이다.

표 3. 블록 크기에 따른 검색 공간 클러스터링 적용 후의 후보 위치 블록의 개수 비교

블록크기 적용여부	20cm	15cm	10cm	5cm	3.3cm
클러스터링 전	4389	9954	33894	266049	882304
클러스터링 후	3256	6290	14398	52159	104119
감소율	24%	37%	58%	80%	88%

표 4. 블록 크기에 따른 검색 공간 클러스터링 적용 후의 연산 시간 비교 (단위 : milli-second)

블록크기 적용여부	20cm	15cm	10cm	5cm	3.3cm
클러스터링 전	5.5	7.8	17.2	104.7	3457
클러스터링 후	5.5	7	9.4	24.2	44.5
감소율	0%	11%	45%	77%	99%

구간 1, 2에서 15cm 크기의 블록으로 검색 공간 클러스터링을 적용한 후보 위치 블록의 개수는 6290개이고, 이를 모든 구간으로 확장하면 8배인 약 50320개로 구간 1, 2에서 5cm 크기의 블록으로 검색 공간 클러스터링을 적용한 후보 위치 블록의 개수인 52159와 유사하다. 이 때 연산 시간은 윈도우 크기인 32ms보다 작은 약 24.2ms로 15cm와 20cm 크기의 블록에 공간 검색 클러스터링을 적용하면 실시간으로 3차원의 모든 구간을 연산할 수 있음을 유추할 수 있다.

IV. 결론

본 논문에서는 3차원 공간에서의 음향 위치 추적을 위해 SRP 기반 방법을 살펴보고, 계산량을 줄이기 위한 검색 공간 클러스터링 방법을 제안하였다. 음향이 위치할 수 있는 후보 위치 블록의 크기에 따라 검색 공간 클러스터링 적용 후의 후보 위치 블록의 개수와 연산 시간을 비교하는 실험을 하였다. 실험 결과 검색 공간 클러스터링을 적용함으로써 후보 위치 블록의

개수와 연산 시간이 감소하는 것과 블록의 크기가 작아질수록 그 감소율이 커지는 것을 알 수 있었다. 또한 3차원 모든 공간에서 SRP 기반 방법을 이용하여 실시간으로 연산 가능한 블록의 크기를 유추할 수 있었다. 그렇지만 블록의 크기가 작아질수록 위치 추적의 오차가 적어지기 때문에 향후, 보다 작은 블록의 크기에서 실시간 연산이 가능한 검색 공간 감소 방법의 연구가 필요하다.

참고문헌

- [1] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing", in *Proc. ICASSP*, vol. 1, pp. 187-190, 1997.
- [2] D. Giuliani, M. Omologo, and P. Svaizer, "Talker localization and speech recognition using a microphone array and a cross-powerspectrum phase analysis", in *Proc. ICSLP*, vol. 22-1, pp. 1243-1246, 1994.
- [3] E. Mumolo, M. Nolich, and G. Vercelli, "Algorithms for acoustic localization based on microphone array in service robotics", *J. Robot. Autom.*, vol. 42, no. 2, pp. 69-88, 2003.
- [4] J. DiBiase, "A high accuracy, low-latency technique for talker localization in reverberant environments", PhD Thesis, Brown University, 2000.
- [5] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276-280, 1986.
- [6] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location", *IEEE Tran. Acoust., Speech, Signal Processing*, vol. 5, no. 3, 1997.
- [7] A. Johansson and S. Nordholm, "Robust acoustic direction of arrival estimation using root-SRP-PHAT, a realtime implementation", in *Proc. ICASSP*, pp. 933-936, 2005.
- [8] D. Zotkin and R. Duraiswami, "Accelerated speech source localization via hierarchical search of steered response power", *IEEE Tran. Acoust., Speech, Signal Processing*, vol. 12, no. 5, pp. 499-508, 2002.