



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0144822
(43) 공개일자 2023년10월17일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G01S 5/20 (2006.01)
G06F 17/15 (2006.01) G06N 3/04 (2023.01)

(52) CPC특허분류
G06N 3/08 (2023.01)
G01S 5/20 (2013.01)

(21) 출원번호 10-2022-0043976
(22) 출원일자 2022년04월08일
심사청구일자 2022년04월08일

(71) 출원인
한국표준과학연구원
대전 유성구 가정로 267(가정동, 한국표준과학연구원)

포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)

(72) 발명자
장지호
경기 성남시 분당구 내정로 185 양지마을 청구아파트 211-1001

조완호
대전광역시 서구 청사로 70 누리아파트 109동 808호

(뒷면에 계속)

(74) 대리인
특허법인 피씨알

전체 청구항 수 : 총 9 항

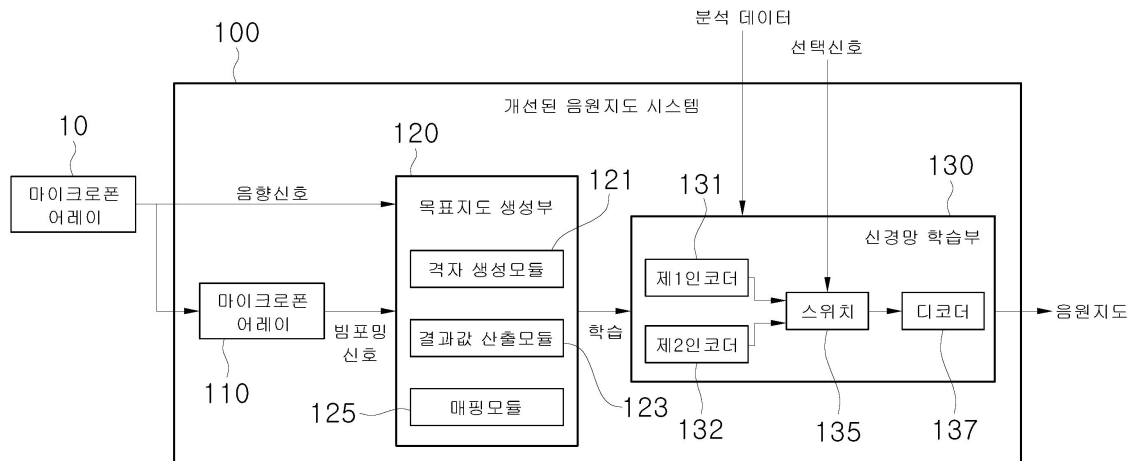
(54) 발명의 명칭 **신경망 학습부 및 이를 포함하는 개선된 음원지도 제공 시스템**

(57) 요약

본 발명은 음원지도 제공 시스템을 개시한다. 보다 상세하게는, 본 발명은 다수의 마이크로폰으로 측정된 음향신호를 이용하여 평면 또는 곡면상에 존재하는 하나 이상의 음원들의 세기 및 위치를 이미지 형태로 표시하는 음원지도를 제공하는 딥러닝 기반의 음원지도 제공 시스템 및 방법에 관한 것이다.

본 발명의 실시예에 따르면, 초음파 마이크로폰 어레이를 이용하여 음향지도를 제공하는 시스템에서, 분석을 위해 도입된 딥러닝 모델에 서로 다른 주파수가 설정된 적어도 두 개 인코더를 탑재함으로써, 저역과 고역을 달리하여 공간 엘리어싱 현상을 최소화할 수 있다.

대표도



- (52) CPC특허분류
G06F 17/153 (2013.01)
G06N 3/048 (2023.01)

이수영

경기도 평택시 비전동 용죽1로 64 비전푸르지오
 105동 1604호

- (72) 발명자
이승철
 경북 포항시 남구 지곡로 155, 4동 1003호

이 발명을 지원한 국가연구개발사업

| | |
|-------------|-------------------------------------|
| 과제고유번호 | 1415174663 |
| 과제번호 | 20206610100290 |
| 부처명 | 산업통상자원부 |
| 과제관리(전문)기관명 | 한국에너지기술평가원 |
| 연구사업명 | 화력발전소안전환경구축기술개발(R&D) |
| 연구과제명 | 발전소 안전환경 정착을 위한 IoT기반 작업안전관리 플랫폼 개발 |
| 기 여 율 | 25/100 |
| 과제수행기관명 | 포항공과대학교 산학협력단 |
| 연구기간 | 2020.05.01 ~ 2024.04.30 |

이 발명을 지원한 국가연구개발사업

| | |
|-------------|-----------------------------|
| 과제고유번호 | 1345344204 |
| 과제번호 | 2020R1A6A1A03047902 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 한국연구재단 |
| 연구사업명 | 기초연구사업 |
| 연구과제명 | 환자 맞춤형 진단 및 치료 시스템 개발 및 사업화 |
| 기 여 율 | 25/100 |
| 과제수행기관명 | 포항공과대학교 산학협력단 |
| 연구기간 | 2020.06.01 ~ 2023.02.28 |

이 발명을 지원한 국가연구개발사업

| | |
|-------------|-------------------------|
| 과제고유번호 | 22011047 |
| 과제번호 | GP2022-0002-12 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 한국표준과학연구원 |
| 연구사업명 | 기관고유사업 |
| 연구과제명 | 국제동등성 확보를 위한 국가측정표준 확립 |
| 기 여 율 | 50/100 |
| 과제수행기관명 | 한국표준과학연구원 |
| 연구기간 | 2022.01.01 ~ 2022.12.31 |

명세서

청구범위

청구항 1

복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제1 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제1 인코더;

복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제2 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제2 인코더;

상기 입력 데이터의 형태에 따라 상기 제1 및 제2 인코더 중 어느 하나의 출력을 선택하는 스위치; 및

상기 스위치의 선택에 따라, 상기 제1 또는 제2 인코더로부터 출력되는 특징 맵을 입력받아 복원하여 출력 이미지를 생성하는 디코더

를 포함하는 신경망 학습부.

청구항 2

제 1 항에 있어서,

상기 제1 및 제2 주파수 대역은,

상기 입력 데이터에 대한 공간 나이퀴스트(Nyquist) 주파수 파장을 기준으로 하여 각각 고대역 및 저대역으로 구분되는 것인, 신경망 학습부.

청구항 3

제 2 항에 있어서,

상기 복수의 계층은,

각각 하나 이상의 컨볼루션 계층, 풀링 계층, 배치 정규화 계층 및 비선형화 계층

을 포함하는 신경망 학습부.

청구항 4

제 3 항에 있어서,

제1 및 제2 컨볼루션 인코더는,

상기 컨볼루션 계층의 커널 크기가 서로 상이하게 설정된 것인, 신경망 학습부.

청구항 5

제 2 항에 있어서,

상기 스위치는,

학습시, 상기 입력 데이터에 공간 엘리어싱이 존재하지 않는 경우 상기 제1 인코더를 선택하고, 상기 입력 데이터에 공간 엘리어싱이 존재하는 경우 상기 제2 인코더를 선택하는 것인, 신경망 학습부.

청구항 6

복수의 마이크로폰이 일정 간격으로 배치되는 마이크로폰 어레이로부터 수신된 음향신호로부터 음원지도를 생성하는 음원지도 제공 시스템으로서,

상기 마이크로폰 어레이로부터 수신한 음향신호로부터 빔형성 과정을 통해 빔포밍 이미지를 생성하는 이미지 생성부;

상기 빔포밍 이미지를 이용하여 목표지도를 생성하는 목표지도 생성부; 및

상기 목표지도를 입력받아 학습을 수행하고, 설정된 딥러닝 모델에 기반하여 이미지 내 포함되는 하나 이상의 주요 특징에 따라 분석 이미지에 포함되는 음원의 위치 및 세기를 이미지화한 음원지도를 생성하는 신경망 학습부

를 포함하는 개선된 음원지도 제공 시스템.

청구항 7

제 6 항에 있어서,

상기 신경망 학습부는,

복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 상기 빔포밍 이미지로부터 제1 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제1 인코더;

복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 상기 빔포밍 이미지로부터 제2 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제2 인코더;

상기 빔포밍 이미지의 형태에 따라 상기 제1 및 제2 인코더 중 어느 하나의 출력을 선택하는 스위치; 및

상기 스위치의 선택에 따라, 상기 제1 또는 제2 인코더로부터 출력되는 특징 맵을 복원하여 상기 음원지도를 생성하는 디코더

를 포함하는 개선된 음원지도 제공 시스템.

청구항 8

제 7 항에 있어서,

상기 제1 및 제2 주파수 대역은,

각각 상기 빔포밍 이미지에 대한 공간 나이퀴스트(Spatial Nyquist) 주파수 파장을 기준으로 하여 고대역 및 저대역으로 구분되는 것인, 개선된 음원지도 제공 시스템.

청구항 9

제 7 항에 있어서,

상기 스위치는,

상기 빔포밍 이미지에 공간 엘리어싱이 존재하지 않는 경우 상기 제1 인코더를 선택하고, 상기 빔포밍 이미지에 공간 엘리어싱이 존재하는 경우 상기 제2 인코더를 선택하는 것인, 개선된 음원지도 제공 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 음원지도 제공 시스템에 관한 것으로, 특히 다수의 마이크로폰으로 측정된 음향신호를 이용하여 평면 또는 곡면상에 존재하는 하나 이상의 음원들의 세기 및 위치를 이미지 형태로 표시하는 음원지도를 제공하는 딥러닝 기반의 음원지도 제공 시스템에 관한 것이다.

배경 기술

[0002] 음원지도란, 위치 및 세기가 알려져 있지 않은 다양한 복수의 음원이 존재할 것으로 여겨지는 평면 또는 곡면 위에, 그 음원들의 위치 및 세기를 계산한 후 그것을 표시하는 이미지를 가리킨다.

[0003] 도 1은 종래의 음원지도를 제공하는 빔 포밍 방법을 모식도로 나타낸 도면으로서, 종래의 음원지도를 생성하는 기술과 관련하여, 빔형성(beamforming) 방법에 의하면, 측정장치의 측정면(Measurement plane)에 다수의 마이크로폰(microphone)을 배열하여 마이크로폰 어레이를 형성하고, 마이크로폰 어레이가 x,y,z 방향에서 수신한 측정

신호로부터 음원면상의 각 음원들의 위치 및 세기를 예측함으로써 빔형성 지도(beamforming map)를 생성할 수 있다.

- [0004] 특히, 이러한 빔형성 방법에서는 음원들의 위치 및 세기는 마이크로폰 어레이를 이용하여 취득한 음압값으로부터 계산하고, 이를 평면상에 매칭하여 지도를 완성하게 된다.
- [0005] 그런데, 전술한 빔형성 방법에 의하면, 음원을 검출하는 마이크로폰의 간격이 반파장보다 커지는 주파수 구간에서는 공간 앨리어싱(spatial aliasing)이 발생하는 문제점이 있다.
- [0006] 전술한 공간 앨리어싱은 음원지도 상의 실제 음원의 위치뿐만 아니라 다른 위치에서도 피크값이 나타남에 따라 실제로 없는 음원이 있는 것처럼 표현되는 현상을 가리킨다. 이러한 공간 앨리어싱은 다른 용어로 유령음원(ghost source) 라고도 한다.
- [0007] 이러한 문제를 해결하기 위해, 인공지능 분야의 딥-러닝(Deep learning) 기술이 적용되었고, 이에 전술한 유령음원 현상을 극복할 수 있는 다양한 솔루션이 제안되고 있다.
- [0008] 한편, 기존의 음원지도 제작방법 중 하나로서, 시간 지연-합 빔형성(delay and sum beamforming) 등에 빔형성 방법에 의한 음원지도의 분해능을 향상시키는 디컨볼루션(deconvolution) 방법들이 있었으나, 이러한 기존 방법을 따르다 하더라도 높은 주파수에서는 여전히 공간 앨리어싱 문제가 발생한다고 알려져 있다. 즉, 일반적인 빔형성 방법뿐 아니라 최신의 딥-러닝 기술을 적용한다 하더라도 공간 앨리어싱의 발생에 의한 유령음원들이 등장 다수 나타나는 문제를 해결하는 데는 한계가 있다.

선행기술문헌

특허문헌

- [0009] (특허문헌 0001) 공개특허공보 제10-2022-0032382호(공개일자: 2022.03.15.)

발명의 내용

해결하려는 과제

- [0010] 본 발명은 전술한 문제점을 해결하기 위해 안출된 것으로, 본 발명은 마이크로폰 어레이를 이용하여 하나 이상의 음원을 검출하고 그 세기 및 위치를 평면상에 매핑한 음원지도를 제공하는 시스템에서 공간 앨리어싱에 의한 음원지도의 정확도 저하를 개선하는데 과제가 있다.
- [0011] 이를 위해, 본 발명은 공간 나이퀴스트 주파수를 기준으로 그보다 낮은 주파수 대역 및 높은 주파수 대역에 대해 딥러닝 모델의 일부를 다르게 구성한 시스템을 제공하는 데 과제가 있다.

과제의 해결 수단

- [0012] 전술한 과제를 해결하기 위해, 본 발명의 바람직한 실시예에 따른 신경망 학습부는, 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제1 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제1 인코더, 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제2 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제2 인코더, 상기 입력 데이터의 형태에 따라 상기 제1 및 제2 인코더 중 어느 하나의 출력을 선택하는 스위치 및, 상기 스위치의 선택에 따라, 상기 제1 또는 제2 인코더로부터 출력되는 특징 맵을 입력받아 복원하여 출력 이미지를 생성하는 디코더를 포함할 수 있다.
- [0013] 상기 제1 및 제2 주파수 대역은, 상기 입력 데이터에 대한 공간 나이퀴스트(Nyquist) 주파수 파장을 기준으로 하여 각각 고대역 및 저대역으로 구분될 수 있다.
- [0014] 상기 복수의 계층은, 각각 하나 이상의 컨볼루션 계층, 풀링 계층, 배치 정규화 계층 및 비선형화 계층을 포함할 수 있다.
- [0015] 제1 및 제2 컨볼루션 인코더는, 상기 컨볼루션 계층의 커널 크기가 서로 상이하게 설정될 수 있다.
- [0016] 상기 스위치는, 학습시, 상기 입력 데이터에 공간 앨리어싱이 존재하지 않는 경우 상기 제1 인코더를 선택하고,

상기 입력 데이터에 공간 앨리어싱이 존재하는 경우 상기 제2 인코더를 선택할 수 있다.

[0017] 또한, 전술한 과제를 해결하기 위해, 본 발명의 바람직한 실시예에 따른 신경망 학습부를 포함하는 개선된 음원 지도 제공 시스템은, 복수의 마이크로폰이 일정 간격으로 배치되는 마이크로폰 어레이로부터 수신된 음향신호로부터 음원지도를 생성하는 음원지도 제공 시스템으로서, 상기 마이크로폰 어레이로부터 수신한 음향신호로부터 빔형성 과정을 통해 빔포밍 이미지를 생성하는 이미지 생성부, 상기 빔포밍 이미지를 이용하여 목표지도를 생성하는 목표지도 생성부 및, 상기 목표지도를 입력받아 학습을 수행하고, 설정된 딥러닝 모델에 기반하여 이미지 내 포함되는 하나 이상의 주요 특징에 따라 분석 이미지에 포함되는 음원의 위치 및 세기를 이미지화한 음원지도를 생성하는 신경망 학습부를 포함할 수 있다.

[0018] 상기 신경망 학습부는, 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 상기 빔포밍 이미지로부터 제1 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제1 인코더, 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 상기 빔포밍 이미지로부터 제2 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제2 인코더, 상기 빔포밍 이미지의 형태에 따라 상기 제1 및 제2 인코더 중 어느 하나의 출력을 선택하는 스위치 및, 상기 스위치의 선택에 따라, 상기 제1 또는 제2 인코더로부터 출력되는 특징 맵을 복원하여 상기 음원지도를 생성하는 디코더를 포함할 수 있다.

[0019] 상기 제1 및 제2 주파수 대역은, 각각 상기 빔포밍 이미지에 대한 공간 나이퀴스트(Spatial Nyquist) 주파수 파장을 기준으로 하여 고대역 및 저대역으로 구분될 수 있다.

[0020] 상기 스위치는, 상기 빔포밍 이미지에 공간 앨리어싱이 존재하지 않는 경우 상기 제1 인코더를 선택하고, 상기 빔포밍 이미지에 공간 앨리어싱이 존재하는 경우 상기 제2 인코더를 선택할 수 있다.

발명의 효과

[0021] 본 발명의 실시예에 따르면, 초음파 마이크로폰 어레이를 이용하여 음향지도를 제공하는 시스템에서, 분석을 위해 도입된 딥러닝 모델에 서로 다른 주파수가 설정된 적어도 두 개 인코더를 탑재함으로써 저역과 고역을 달리 하여 공간 앨리어싱 현상을 최소화할 수 있다.

[0022] 이에 따라, 학습시 딥러닝 모델의 컨볼루션 연산에서 더 넓은 영역을 참조할 수 있게 되어 마이크로폰 어레이의 설계시 센서간 간격을 보다 좁게 설정할 수 있고, 이에 마이크로폰 어레이에 필요한 마이크로폰의 개수를 줄임으로써 시스템 구축비용을 절감할 수 있는 효과가 있다.

도면의 간단한 설명

- [0023] 도 1은 종래의 음원지도를 제공하는 빔 포밍 방법을 모식도로 나타낸 도면이다.
- 도 2는 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 구조를 나타낸 도면이다.
- 도 3은 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템에 의한 목표지도 생성 방법을 모식화한 도면이다.
- 도 4는 본 발명의 실시예에 따른 개선된 음원지도 제공 시스템에 포함되는 신경망 학습부의 네트워크 구조를 나타낸 도면이다.
- 도 5는 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 기계학습 방법을 나타낸 도면이다.
- 도 6은 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 음원지도 제공 방법을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

- [0025] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.
- [0026] 또한, 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어는 본 발명이 속하는 기술분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [0027] 이하의 설명에서, 본 발명의 "신경망 학습부" 및 "개선된 음원 제공 시스템" 및 이를 이루는 각 구성부는, 공지의 마이크로 프로세서에 의해 실행 가능하고, 읽고 쓰기가 가능한 기록매체에 저장되어 컴퓨팅 장치에 탑재될 수 있다.
- [0028] 이하의 설명에서, 본 발명의 "신경망 학습부를 포함하는 개선된 음원지도 제공 시스템"을 가리키는 용어는, 기재의 편의상 "음원지도 제공 시스템" 또는 "시스템"으로 약식 기재될 수 있다.
- [0029] 이하, 도면을 참조하여 본 발명의 실시예에 따른 신경망 학습부 및 이를 포함하는 개선된 음원지도 제공 시스템을 설명한다.
- [0030] 도 2는 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 구조를 나타낸 도면이고, 도 3은 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템에 의한 목표지도 생성방법을 모식화한 도면이다.
- [0031] 도 2를 참조하면, 본 발명의 실시예에 따른 음원지도 제공 시스템(100), 마이크로폰 어레이(10)로부터 수신한 음향신호로부터 빔형성 과정을 통해 빔포밍 이미지를 생성하는 이미지 생성부(110), 빔포밍 이미지 이용하여 목표지도를 생성하는 목표지도 생성부(120) 및, 목표지도를 입력받아 학습을 수행하고, 설정된 딥러닝 모델에 기반하여 이미지 내 포함되는 하나 이상의 주요 특징에 따라 분석 이미지에 포함되는 음원의 위치 및 세기를 이미지화한 음원지도를 생성하는 신경망 학습부(130)를 포함할 수 있다.
- [0032] 상세하게는, 이미지 생성부(110)는 마이크로폰 어레이(10)가 음원으로부터 대한 음향신호를 입력받아 시간지연 합 빔형성(delay-and-sum beamforming) 등의 빔 형성 기술을 통해 빔포밍 이미지를 생성할 수 있다.
- [0033] 여기서, 마이크로폰 어레이(10)는 복수개가 좌우로 배열된 마이크로폰으로 이루어지며, 마이크로폰들이 배열된 전체 너비는 저대역 주파수의 음향신호에 영향을 주고, 각 마이크로폰의 간격은 고대역 주파수의 음향신호에 영향을 준다.
- [0034] 이에, 종래 마이크로폰 어레이(10)의 설계자는 음향신호에서 발생할 수 있는 공간 앨리어싱 현상 등을 고려하여 마이크로폰 어레이(10)의 마이크로폰간 간격을 좁게 배치하는 등, 어레이 구성을 다르게 설계하여야 하나, 본 발명의 개선된 음지도 제공 시스템(100)에 의하면, 음향신호에 따라 적절한 인코더를 통해 음원지도를 생성함으로써, 마이크로폰 어레이의 설계를 변경하지 않고도 저대역 및 고대역 주파수 모두를 포괄할 수 있는 것을 특징으로 한다.
- [0035] 목표지도 생성부(120)는 음향신호에 기초하여 학습 데이터로 이용되는 목표지도를 생성하고 신경망 학습부(130)에 입력하여 딥 러닝이 수행되도록 한다. 이를 위해, 목표지도 생성부(120)는 그 기능에 따라 복수의 모듈로 구성될 수 있다.
- [0036] 예를 들면, 목표지도 생성부(120)는 음원의 주변에 소정 범위의 간격을 갖는 격자를 생성하는 격자 생성모듈(121), 음원의 위치에서 결과값이 최대가 되고, 음원과 거리 따라 결과값이 감소하도록 격자의 각 좌표별 결과값을 산출하는 결과값 산출모듈(123), 결과값을 격자의 각 좌표와 대응되는 행렬 상의 위치에 배열하고, 그 행렬에 배열된 결과값을 이용하여 이미지 형태의 목표지도를 생성하는 매핑모듈(125)을 포함할 수 있다.
- [0037] 도 3을 함께 참조하면, 격자 생성모듈(121)은 하나 이상의 음원이 존재하는 특정 영역에 대하여 일정한 간격을 갖는 격자점을 포함하는 격자를 생성할 수 있고, 각 음원의 위치를 격자상에 표시할 수 있다.

- [0038] 여기서, 격자를 이루는 격자점간 간격(K)이 좁을수록 더욱 정확한 음원지도의 취득이 가능하나 연산시간은 길어지고, 간격(K)이 넓을수록 음원지도를 취득하는데 걸리는 연산시간은 짧아지며, 음원지도의 정확도는 다소 낮아지게 된다. 이에 따라, 본 발명의 실시예에 따르면, 격자 생성모듈(121)이 제공하는 격자의 바람직한 격자점간 간격(K)은 반복적인 실험을 통해 적절한 크기로 설정될 수 있다.
- [0039] 결과값 산출모듈(123)은 격자 생성모듈(121)이 제공하는 격자의 격자점들 $\{(1,1), (1,2) \sim (N,M); N,M \text{은 자연수}\}$ 과 격자상에 위치하는 하나 이상의 음원(A,B,C)과의 거리(R)로부터 결과값을 산출할 수 있다.
- [0040] 여기서, 결과값은 취득한 음향신호를 이미지화 할 때 각 픽셀(Pixel)이 갖는 특정 값을 의미하며, 음원(A,B,C)과 격자의 각 좌표 사이의 거리(R)를 이용하여 산출될 수 있다.
- [0041] 또한, 결과값 산출에 있어서, R의 승수를 통해 그 조절함으로써 거리(R)에 따른 결과값의 변화율을 조절할 수 있다. 즉, R의 제곱근의 승수를 상대적으로 큰 값으로 설정함으로써 거리(R)에 따른 결과값의 변화율을 크게 하여, 음원지도 내 음원의 위치나 세기의 차이를 더욱 명확하게 나타낼 수 있다.
- [0042] 맵핑모듈(125)은 전술한 거리(R)를 이용하여 산출해낸 결과값을 격자에 대응되는 행렬상의 위치에 배열할 수 있다. 특히, 맵핑모듈(125)은 모든 좌표에 대한 결과값을 산출할 때까지 작업을 반복하여 격자 생성모듈(121)이 생성한 격자와 동일한 크기의 행렬을 생성할 수 있고, 그 생성한 행렬에 결과값을 배열하고, 이를 이용하여 이미지 형태의 목표지도를 생성할 수 있다.
- [0043] 이에 따라, 목표지도 생성부(120)를 통해 생성한 목표지도는 후술하는 신경망 학습부(130)에 입력되어 해당 빔포밍 이미지에 대한 학습 데이터로서 이용될 수 있다.
- [0044] 신경망 학습부(130)는 목표지도 생성부(120)로부터 목표지도를 입력받아 기계학습을 반복 수행하여 다양한 음원 즉, 빔포밍 이미지에 대한 특징 맵을 구축함과 더불어, 다양한 주파수 대역의 음향신호에 대응하는 빔포밍 이미지 형태의 분석 데이터를 입력받아 특징을 추출함으로써 공간 엘리어싱 현상이 제거된 데이터를 추출하고 음원 지도를 생성 및 제공할 수 있다.
- [0045] 공간 엘리어싱 현상은 음원지도의 전반에 퍼져서 나타남에 따라 딥러닝 모델의 구조가 공간 나이퀴스트 주파수보다 낮은 주파수에서의 구조가 서로 달라져야 할 필요가 있다. 여기서, 공간 나이퀴스트 주파수는 마이크로폰 어레이 상의 이웃한 마이크로폰간 간격이 반파장과 같아지는 주파수를 가리킨다.
- [0046] 이를 위해, 본 발명의 실시예에 따르면, 시스템에 적용된 딥러닝 모델은 컨볼루션 인코더-디코더(convolutional encoder-decoder) 네트워크를 기반으로 한 2인코더-1디코더 구조일 수 있다.
- [0047] 상세하게는, 본 발명의 실시예에 따른 신경망 학습부(130)는, 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제1 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제1 인코더(131), 복수의 계층으로 이루어지고, 컨볼루션 연산을 통해 입력 데이터로부터 제2 주파수 대역에서 추출되는 하나 이상의 주요 특징을 포함하는 특징 맵을 생성하는 제2 인코더(132), 입력 데이터의 형태에 따라 상기 제1 및 제2 인코더(131, 132) 중 어느 하나의 출력을 선택하는 스위치(135) 및, 스위치(135)의 선택에 따라, 제1 또는 제2 인코더(131, 132)로부터 출력되는 특징 맵을 입력받아 복원하여 출력 이미지를 생성하는 디코더(137)를 포함할 수 있다.
- [0048] 제1 인코더(131) 및 제2 인코더(132)는 복수의 컨볼루션 계층으로 이루어지는 인코더로서, 입력 데이터에 대하여 서로 다른 주파수 기준에 따라 특징 맵을 추출할 수 있다. 일례로서, 제1 인코더(131)는 공간 엘리어싱이 발생하지 않은 일반적인 주파수 파장을 갖는 입력 데이터에 적합하도록 설정되어 있고, 제2 인코더(132)는 공간 엘리어싱이 발생하는 주파수 파장을 갖는 입력 데이터에 적합하도록 설정될 수 있다.
- [0049] 여기서, 전술한 주파수 기준은 공간 나이퀴스트 주파수를 기준으로 설정되되, 마이크로폰 어레이의 크기 및 마이크로폰간 거리에 의해 다르게 설정될 수 있다.
- [0050] 이에 따라, 제1 인코더(131)는 일반적인 주파수 파장을 갖는 입력 데이터로부터 특징 맵을 생성하도록 설계될 수 있고, 제2 인코더(132)는 이보다 높은 주파수 파장을 갖는 입력 데이터로부터 특징맵을 생성하도록 설계될 수 있다.
- [0051] 이러한 제1 및 제2 인코더(131, 132)를 포함하여 디코더(137)에 관한 구체적 네트워크 구조 및 그 특징에 대한 상세한 설명은 후술한다.
- [0052] 스위치(135)는 제1 및 제2 인코더(131, 132) 중, 선택되는 어느 하나와 후술하는 디코더(137)를 서로 연결하는

역할을 한다. 특히, 본 발명의 실시예에 따르면, 분석 데이터의 분석시 주파수 파장에 따라 서로 다른 설정값을 갖는 하나의 인코더를 선택적으로 운영하는 것으로, 스위치(135)는 분석 데이터에 대한 파장을 통해 어떤 주파수 파장을 갖는지 판단하고, 어느 하나의 인코더에서 생성한 특징 맵을 디코더(137)에 전달할 수 있다.

[0053] 디코더(137)는 인코더에 의해 추출된 특징 맵을 디코딩을 통해 복원하여 출력 데이터를 생성할 수 있다. 디코더(137)는 인코더와 대칭되는 구조로 구성될 수 있고, 특징 맵에 대한 디컨볼루션 과정을 통해 특정 주파수 성분 에 대한 특징이 극대화된 이미지를 도출할 수 있다. 이에 따라 분역된 분석 데이터가 하나 이상의 음원에 대한 빔포밍 이미지인 경우, 그 음원의 위치 및 세기가 극대화된 음원지도도를 제공하게 된다.

[0054] 전술한 구조에 따라, 본 발명의 실시예에 따른 개선된 음원지도 시스템(100)은 딥 러닝 기술을 통해 빔포밍 이미지로부터 음원의 위치 및 세기를 직관적으로 확인할 수 있는 음원지도도를 제공할 수 있으며, 특히 신경망 학습부를 서로 다른 주파수 파장에 따라 다른 특징 맵을 추출하도록 설정된 적어도 두 개의 인코더로 구축함으로써 공간 엘리어싱 현상이 최소화된 음원지도도를 제공할 수 있다.

[0055] 이하, 도면을 참조하여 본 발명의 실시예에 따른 개선된 음원지도 제공 시스템에 탑재되는 신경망 학습부의 네트워크 구조를 통해 본 발명의 기술적 사상을 상세히 설명한다.

[0056] 도 4는 본 발명의 실시예에 따른 개선된 음원지도 제공 시스템에 포함되는 신경망 학습부의 네트워크 구조를 나타낸 도면이다.

[0057] 도 4를 참조하면, 본 발명의 실시예에 따른 신경망 학습부는, 두 개의 컨볼루션 인코더 및 한 개의 컨볼루션 디코더로 구성될 수 있다. 각 인코더와 디코더는 상호 대칭되는 구조로서, 컨볼루션 인코더-디코더는 입력 이미지에서 추출되는 특징 맵(feature map)의 크기를 줄였다가 다시 입력 이미지의 크기만큼 크게 만들어서 입력 이미지의 각 픽셀에 대해 분류 결과로 클래스를 분류할 수 있다.

[0058] 두 개의 컨볼루션 인코더는 각각 복수의 계층(layer)으로 구성될 수 있다. 복수의 계층은 적어도 하나의 컨볼루션 계층(convolutional layer)과, 풀링 계층(pooling layer)과, 배치 정규화 계층(batch normalization layer)과, 비선형화 계층(non linear activation layer)을 포함할 수 있다. 그리고 컨볼루션 인코더에서는 전술한 각 계층이 순차적으로, 컨볼루션 계층, 배치 표준화 계층 및 비선형화 계층으로 배열될 수 있다.

[0059] 또한, 하나의 계층은 채널 수(ch) 만큼 반복될 수 있다. 즉 하나의 계층은 '(컨볼루션 계층, 배치 표준화 계층 및 비선형화 계층) × 채널 수(ch) + 풀링 계층' 구조를 가질 수 있다.

[0060] 이러한 컨볼루션 계층은 입력 이미지에 대한 컨볼루션 연산을 통해 특징맵(feature map)을 출력할 수 있다. 이때, 컨볼루션 연산을 수행하는 필터(filter)를 커널(kernel)이라고 한다. 또한, 커널을 구성하는 연산 파라미터(parameter)를 커널 파라미터(kernel parameter) 또는 가중치(weight)라고 한다. 컨볼루션 계층에서는 하나의 입력에 서로 다른 종류의 커널을 사용할 수 있다.

[0061] 그리고, 컨볼루션 계층은 입력 이미지의 특정 영역을 대상으로 컨볼루션 연산을 수행하며, 그 연산 영역을 윈도우(window)라고 한다. 이러한 윈도우는 영상의 좌측 상단에서 우측 하단까지 한 칸씩 이동할 수 있고, 한 번에 이동하는 이동 크기를 조절할 수 있다. 이때, 이동 크기를 스트라이드(stride)라고 한다.

[0062] 컨볼루션 계층은 입력 이미지에서 윈도우를 이동하면서 입력 이미지의 모든 영역에 대하여 컨볼루션 연산을 수행하게 된다.

[0063] 일례로서, 입력 값에 컨볼루션 연산을 수행한다고 할 때, 입력 크기(N, C_{in}, H, W) 및 출력(N, C_{out}, H, W)에 대한 컨볼루션 계층의 출력값은 이하의 수학적 식 1로 표현할 수 있다(단, 'N'은 배치 크기, 'C'는 채널의 수, 'H'는 커널의 높이, 'W'는 커널의 폭).

수학적 식 1

[0064]
$$out(N_i, C_{out_j}) = bias(C_{out_j}) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) * input(N_i, k)$$

- [0066] 여기서, 'out'은 출력 데이터, 'bias'는 편향값, 'weight'는 가중치, '*'는 컨볼루션, 'input'은 입력 데이터를 나타낸다.
- [0067] 이에, 본 발명의 실시예에 따르면, 제1 및 제2 컨볼루션 인코더는 각각 서로 다른 커널 크기 $\{(H,W)=k\}$ 즉, 'k'값에 따라 컨볼루션 과정을 수행하게 된다.
- [0068] 특히, 본 발명의 실시예에 따르면, 두 개의 컨볼루션 인코더는 커널 크기(k)가 각각 입력 데이터의 주파수 대역에 대응하여 서로 다르게 설정됨에 따라 서로 다른 특징 맵을 생성하는 것을 특징으로 한다.
- [0069] 컨볼루션 인코더는 컨볼루션 과정에서 커널 크기(k)에 따라 추출되는 특징 맵이 결정한다. 이러한 컨볼루션 과정에서 상대적으로 작은 크기의 커널은 저대역 주파수의 이미지, 즉 공간 앨리어싱이 발생하지 않는 이미지에 적합하고, 상대적으로 큰 크기의 커널은 고대역 주파수 이미지, 즉 공간 앨리어싱이 발생함에 따라 이를 최소화해야 하는 이미지에 적합하다.
- [0070] 따라서, 본 발명의 제1 컨볼루션 인코더는 공간 나이퀴스트 주파수를 기준으로 하여 상대적으로 커널 크기(k)가 작은 값(ex. k=5)으로 설정되고, 제2 컨볼루션 인코더는 상대적으로 커널 크기(k)가 큰 값(ex. k=7)을 설정될 수 있다.
- [0071] 풀링 계층은 컨볼루션 계층의 연산 결과로 얻은 특징 맵을 서브 샘플링(sub sampling)한다. 풀링 연산에는 최대 풀링(max pooling)과 평균 풀링(average pooling) 등이 있다. 최대 풀링은 윈도우 내에서 가장 큰 샘플 값을 택한다. 평균 풀링은 윈도우에 포함된 값의 평균 값으로 샘플링한다. 일반적으로 풀링은 스트라이드와 윈도우의 크기가 갖도록 하는 것이 일반적이다.
- [0072] 비선형화 계층은 뉴런에서 출력값을 결정하는 계층으로서, 소정의 전달 함수(transfer function)를 사용할 수 있다. 알려진 전달 함수로는 Relu, sigmoid 등이 있고, 본 발명의 실시예에 따른 시스템에서는 Relu 함수를 사용할 수 있다.
- [0073] 진술한 계층 구조에 따라, 두 개의 컨볼루션 인코더는 입력 이미지에 대한 특징 맵을 생성할 수 있다.
- [0074] 스위치는 두 컨볼루션 인코더 중, 어느 하나의 출력을 후술하는 컨볼루션 디코더에 전달할 수 있다. 이러한 스위치는 입력 데이터의 주파수 대역에 따라 제어되며, 어느 하나의 컨볼루션 인코더의 출력을 컨볼루션 디코더에 전달하게 된다.
- [0075] 컨볼루션 디코더는 두 컨볼루션 인코더가 생성한 특징 맵을 이용하여 출력 이미지를 생성할 수 있다. 두 컨볼루션 인코더에 대비하여 컨볼루션 디코더는 하나가 탑재되며, 복수의 계층으로 구성될 수 있다.
- [0076] 하나의 계층은 업샘플링 계층(upsampling layer) 및 디컨볼루션 계층(deconvolutional layer)을 포함할 수 있다. 디컨볼루션 계층은 컨볼루션 인코더의 컨볼루션 계층의 역동작을 수행할 수 있다. 또한, 디컨볼루션 계층은 컨볼루션 계층, 배치 표준화 계층, 비 선형화 계층의 구조가 반복될 수 있다.
- [0077] 여기서, 컨볼루션 디코더는 컨볼루션 인코더와 대칭적인 구조를 가질 수 있음에 따라, 컨볼루션 인코더와 유사한 개수의 컨볼루션 계층, 동일한 컨볼루션 커널 크기(k) 및 개수를 갖는다.
- [0078] 업샘플링 계층은 풀링 계층의 역동작을 수행할 수 있다. 업샘플링 계층은 업샘플링(upsampling)을 진행할 수 있으며, 풀링 계층은 역으로 차원을 확대하는 역할을 한다.
- [0079] 디컨볼루션 계층은 컨볼루션 계층의 역동작을 수행할 수 있다. 디컨볼루션 계층은 컨볼루션 계층과 반대 방향으로 컨볼루션 연산을 수행하는 것으로, 디컨볼루션 계층은 입력으로 특징 맵을 받아 커널을 이용한 컨볼루션 연산으로 시각화를 위한 출력 이미지를 생성할 수 있다.
- [0080] 여기서, 스트라이드를 1로 하면 디컨볼루션 계층은 특징맵의 가로, 세로 크기가 출력의 가로, 세로와 동일한 이미지를 출력할 수 있고, 스트라이드를 2로 하면 디컨볼루션 계층은 특징 맵의 가로, 세로 크기 대비 절반 크기의 영상을 출력할 수 있다.
- [0081] 이에 따라, 디컨볼루션 절차를 모두 완료한 출력 이미지는 선택된 인코더로부터 공간 앨리어싱이 최소화된 특징 맵에 의해 복원됨에 따라 일정 수준 이상의 높은 정확도를 갖게 된다.
- [0082] 이하, 도면을 참조하여 본 발명의 실시예에 따른 시스템에 의한 음원지도 생성방법을 상세히 설명한다.
- [0083] 도 5는 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 기계학습 방법을 나

타넨 도면이고, 도 6은 본 발명의 실시예에 따른 신경망 학습부를 포함하는 개선된 음원지도 제공 시스템의 음원지도 제공 방법을 나타낸 도면이다.

- [0084] 이하의 설명에서, 별도의 기재가 없는 경우 각 단계별 실행주체는 본 발명의 실시예에 따른 신경망 학습부, 시스템 및 이의 각 구성부가 된다.
- [0085] 도 5를 참조하면, 본 발명의 실시예에 따른 개선된 음원지도 제공 시스템을 통한 기계학습 방법에 따르면, 데이터 수집을 위해 하나 이상의 마이크로폰이 배열된 마이크로폰 어레이가 음원으로부터 음향신호를 수신하면(S100), 시스템은 마이크로폰 어레이로부터 음향신호를 전기신호 형태로 입력받게 된다.
- [0086] 또한, 전술한 S100 단계는 음원지도의 생성을 위해 실시간으로 음향을 수집하는 방식 뿐만 아니라, 별도의 마이크로폰 어레이를 통해 미리 수신하여 녹음된 음향을 수신하는 방식도 적용될 수 있다.
- [0087] 다음으로, 시스템은 입력된 음향신호를 이를 이용하여 빔포밍 기술을 통해 빔포밍 이미지를 생성한다(S110). S110 단계에는 시간지연 합 빔형성(delay-and-sum beamforming) 등의 빔 형성 기술들이 이용될 수 있다.
- [0088] 그리고, 시스템은 빔포밍학습을 위한 학습 데이터를 확보하기 위해 목표지도를 생성하는 절차를 수행할 수 있다.
- [0089] 다음으로, 시스템은 음향신호로부터 하나 이상의 음원이 존재하는 특정 영역에 대하여 일정한 간격을 갖는 격자점을 포함하는 격자를 생성한다(S120). 이때, 각 음원의 위치를 격자상에 표시되며, 격자의 바람직한 격자점간의 간격은 반복적인 실험을 통해 적절한 크기로 설정된다.
- [0090] 이어서, 시스템은 격자점들과 격자상에 위치하는 하나 이상의 음원의 위치 및 거리에 따라 음향신호의 이미지화시, 픽셀이 갖는 특정값인 결과값을 산출한다(S130).
- [0091] 이어서, 시스템은 좌표상에서 음원과의 거리를 이용하여 산출해낸 결과값을 격자에 대응되는 행렬상의 위치에 배열하고 이를 이용하여 이미지 형태의 목표지도를 생성한다(S140).
- [0092] 이후, 시스템은 탑재된 신경망 학습부에 목표지도를 입력하여 해당 빔포밍 이미지에 대한 학습 데이터로서 학습을 수행한다(S150).
- [0093] 다음으로, 도 6을 참조하면 본 발명의 실시예에 따른 시스템을 이용하여 음원지도를 제공하는 방법에 의하면, 마이크로폰 어레이로부터 획득한 음향신호를 빔포밍 이미지 형태로 변환하고, 이를 학습된 신경망 학습부에 입력하여 컨볼루션을 통한 특징 추출을 통해 음원의 위치 및 세기가 이미지상에 명확하게 표시되는 음원지도를 생성할 수 있다.
- [0094] 이를 위해, 신경망 학습부는 분석 데이터를 입력받으면, 기준 이하 또는 미만의 크기의 커널이 설정된 제1 인코더를 통해 분석 데이터로부터 저대역 주파수의 주요특징을 추출하고 이를 포함하는 특징 맵을 생성한다(S200). 또한, 신경망 학습부는 분석 데이터에 대하여 기준 초과 또는 이상의 크기의 커널이 설정된 제2 인코더를 통해 분석 데이터로부터 고대역 주파수의 주요특징을 추출하고 이를 포함하는 특징 맵을 생성한다(S210).
- [0095] 다음으로, 신경망 학습부는 분석 데이터의 형태, 즉 분석 데이터의 주파수 파장이 전술한 마이크로폰간 간격이 반파장 보다 커지는 주파수인 경우 제2 인코더를 선택하고, 그렇지 않은 경우 제1 인코더를 선택한다(S220).
- [0096] 다음으로, S220 단계에서 선택된 인코더로부터 출력되는 특징 맵은 디코더에 달되고, 디코더는 특징 맵을 디컨볼루션 절차를 통해 복원하여 음원의 세기 및 위치가 정확하게 표시되며, 고대역 주파수일 경우 발생하는 공간 앨리어싱이 최소화된 음원지도를 생성한다(S230).
- [0097] 상기한 설명에 많은 사항이 구체적으로 기재되어 있으나 이것은 발명의 범위를 한정하는 것이라기보다 바람직한 실시예의 예시로서 해석되어야 한다. 따라서, 발명은 설명된 실시예에 의하여 정할 것이 아니고 특허청구범위와 특허청구범위에 균등한 것에 의하여 정하여져야 한다.

부호의 설명

- [0098] 10 : 마이크로폰 어레이 100 : 개선된 음원지도 시스템
- 110 : 이미지 생성부 120 : 목표지도 생성부
- 121 : 격자 생성모듈 123 : 결과값 산출모듈

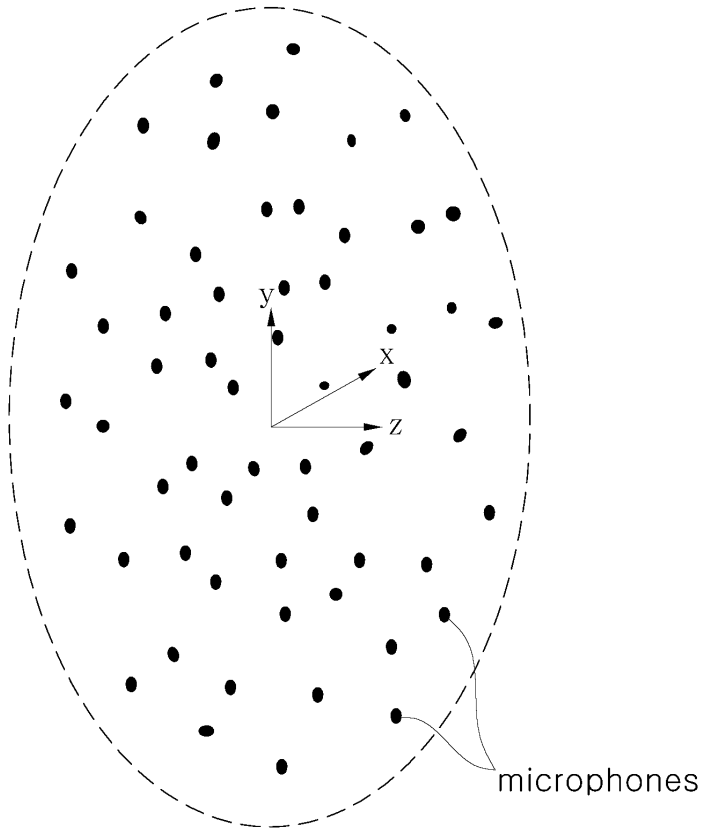
125 : 매핑모듈 130 : 신경망 학습부

131 : 제1 인코더 132 : 제2 인코더

135 : 스위치 137 : 디코더

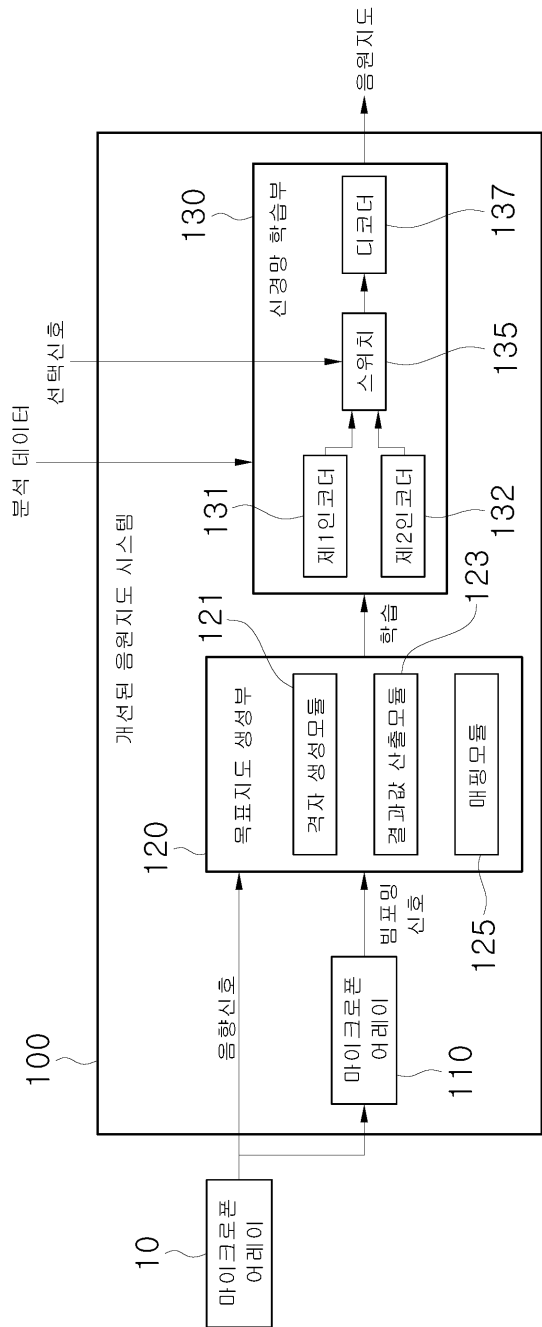
도면

도면1

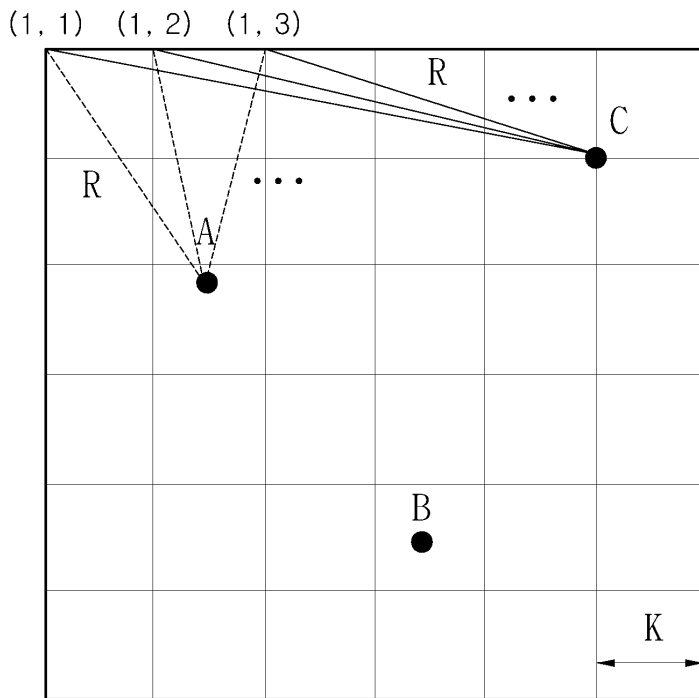


Measurement plane
(z-z_σ)

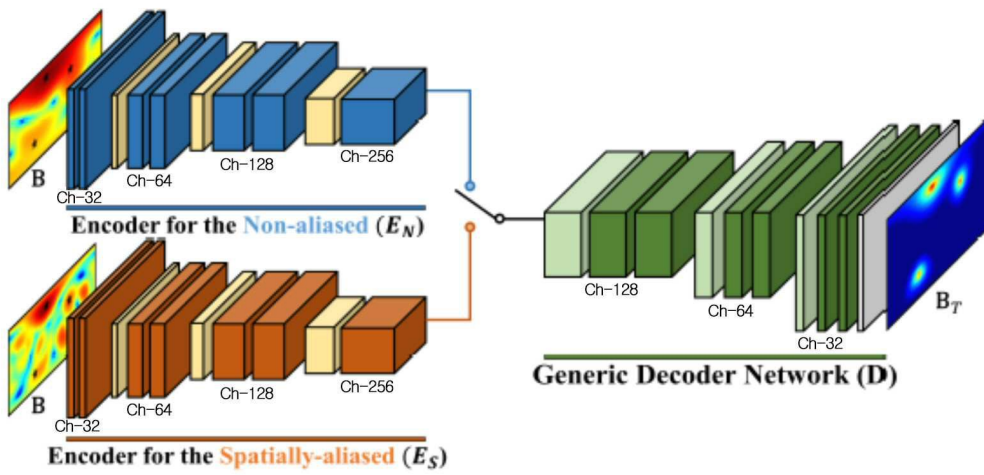
도면2



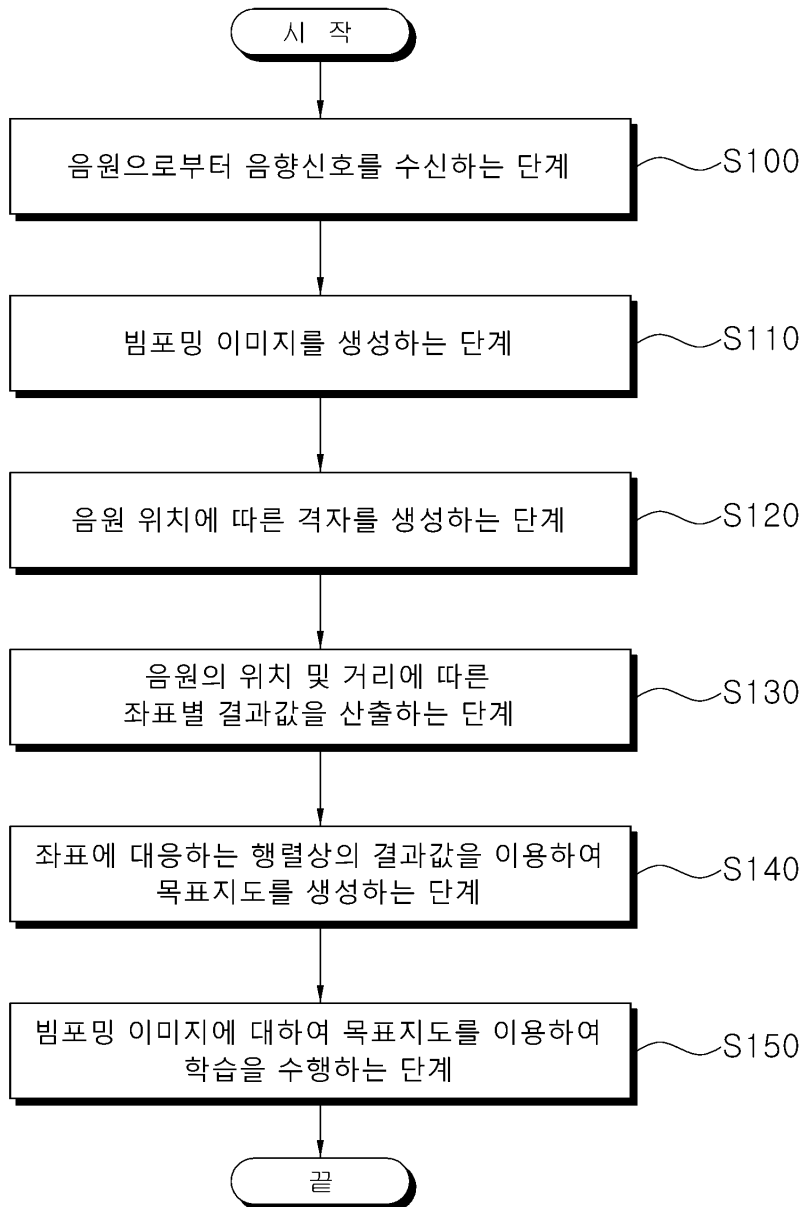
도면3



도면4



도면5



도면6

