

<https://doi.org/10.7236/JIIBC.2021.21.3.59>

JIIBC 2021-3-9

실내 환경에서 Chirp Emission과 Echo Signal을 이용한 심층신경망 기반 객체 감지 기법

DECODE: A Novel Method of DEep CNN-based Object DEtection using Chirps Emission and Echo Signals in Indoor Environment

남현수*, 정종필**

Hyunsoo Nam*, Jongpil Jeong**

요약 인간은 오감 (시각, 청각, 후각, 촉각, 미각) 중 시각 및 청각 정보를 위주로 사용하여 주변 물체를 인식한다. 최신의 객체 인식과 관련한 주요 연구에서는 주로 이미지센서 정보를 이용한 분석에 초점이 맞추어져 있다. 본 논문에서는 다양한 chirp 오디오 신호를 관측공간에 방출하고 2채널 수신센서를 통해 echo를 수집하여 스펙트럼 이미지로 변화시킨 후 딥러닝을 기반으로 이미지 학습 알고리즘을 이용하여 3D 공간상의 객체 인식 실험을 진행하였다. 본 실험은 무향실의 이상적 조건이 아닌 일반적인 실내 환경에서 발생하는 잡음 및 echo가 있는 환경에서 실험을 진행하였고 echo를 통해 객체 인식률을 83% 정확도로 물체의 위치 추정할 수 있었다. 또 한 추론 결과를 관측공간과 3D Sound 공간 신호로 mapping 하여 소리로 출력하여 3D 사운드의 학습을 통해 소리를 통한 시각 정보를 얻을 수 있었다. 이는 객체 인식 연구를 위해서 이미지 정보와 함께 다양한 echo 정보의 활용이 요구된다는 의미이며 이런 기술을 3D 사운드를 통한 증강현실 등에 활용 가능할 것이다.

Abstract Humans mainly recognize surrounding objects using visual and auditory information among the five senses (sight, hearing, smell, touch, taste). Major research related to the latest object recognition mainly focuses on analysis using image sensor information. In this paper, after emitting various chirp audio signals into the observation space, collecting echoes through a 2-channel receiving sensor, converting them into spectral images, an object recognition experiment in 3D space was conducted using an image learning algorithm based on deep learning. Through this experiment, the experiment was conducted in a situation where there is noise and echo generated in a general indoor environment, not in the ideal condition of an anechoic room, and the object recognition through echo was able to estimate the position of the object with 83% accuracy. In addition, it was possible to obtain visual information through sound through learning of 3D sound by mapping the inference result to the observation space and the 3D sound spatial signal and outputting it as sound. This means that the use of various echo information along with image information is required for object recognition research, and it is thought that this technology can be used for augmented reality through 3D sound.

Key Words : Audio-Visual, Chirp, echolocator, Deep convolutional neural network (DCNN)

*준회원, 성균관대학교 스마트팩토리융합학과 석사과정

**정회원, 성균관대학교 스마트팩토리융합학과(교신저자)

접수일자 2020년 12월 31일, 수정완료 2021년 4월 30일

계재확정일자 2021년 6월 4일

Received: 31 December, 2020 / Revised: 30 April, 2021 /

Accepted: 4 June, 2021

**Corresponding Author: jpjeong@skku.edu

Department of Smart Factory Convergence, Sungkyunkwan University, Korea

1. 서 론

최근 컴퓨터 비전^[1]을 이용한 객체 인식은 고전적인 통계적 확률에 기반 한 기계학습 형태^[2]에서 딥러닝 기반 인공지능 알고리즘 활용으로 변화하고 있으며 딥러닝 알고리즘의 성능이 개선된 모델들의 지속적인 등장과 멀티코어 아키텍처의 GPU의 급속한 성능 향상과 가격하락에 따라 컴퓨터 비전 문제 해결에 적용 가능한 범위가 확대되고 있으며 그 결과 다양한 분야에서 인간의 정확성을 뛰어넘는 결과를 얻기 위해 비용대비 효과가 수용 가능한 범위로 들어왔으며 실생활과 다양한 산업현장에서 활발히 적용되어 사용되고 있다. 컴퓨터 비전을 통한 객체의 인식은 기본적으로 배경과 객체의 분리^[3-5] 필요하며 분리된 객체를^[6] 인식해야 하며 객체의 외형을 따라 바운딩 박스^[3]를 추출할 수 있어야 하며 객체의 상태를 설명 할 수 있어야 한다^[7]. 또한 객체 관측 시 획득한 이미지가 완벽하지 않거나 주변 상황변화에 따라 예를 들어 날씨에 따른 빛의 변화, 역광, 반사, 안개, 비, 먼지, 눈 등에 의해 획득 이미지의 해상도가 저하 되는 상황에서도 허용 가능한 오류 범위 내에서의 결과 값을 출력하는 환경 변화에 대한 강건성(Robustness)이 요구 된다.

이 중에서 Indoor 환경에서 객체의 인식과 객체 배경 분리 데이터획득을 위해서 LiDAR, RGBD camera, IR camera 등의 광학센서를 이용하고 있으며 여러 개의 센서 데이터를 이용한 센서 퓨전에 의해서 물체의 감지 능력을 개선하고 있다. 자율주행 차량 등에서는 정보의 다양성과 시스템 안전성을 위해 센서 퓨전 처리가 필수적인 요구사항이다. 빛을 기반으로 하는 센서 모듈을 여러 개 사용하면 정확도 면에서는 이점이 있지만, 센서 모듈의 비용이 증가하며 또한 처리해야 하는 데이터의 양이 증가하기 때문에 의미 있는 응답속도를 위해서는 고성능의 연산 속도를 가지는 프로세스를 필요하게 된다. echo 데이터는 영상 대비 적은 양의 데이터로 물체의 위치와 형태를 감지에 활용될 수 있다^[7].

인간의 시각 정보 처리는 주로 전방을 향하고 있으며 중심축으로부터 멀어지면 이미지의 상세한 정보를 처리하지 않으며 중앙으로 집중된 처리를 하고 있으며 측면 또는 후면의 물체에 인식에 대해서는 소리 신호에 의존해서 위험한 물체의 존재, 위치, 거리 등을 감지한다. 예를 들면 빠르게 접근하는 물체의 감지는 거리에 따라 변하는 소리의 주파수 차이에 의해 물체의 위치 인식한다. 시각장애인의 네비게이션 및 자율주행 로봇 및 무인 운반차 (Automated Guided Vehicle)에서 이런 소리 정

보를 활용하면 저비용으로 광범위한 영역의 물체 감지 효율을 증대시킬 수 있을 것이다.

소리를 이용해서 물체를 감지하는 동물 박쥐, 돌고래의 경우 특정 형태의 소리를 방출하여 echo 신호를 수신하여 어두운 야간 또는 시야가 좋지 않은 수중에서도 물체를 감지하며 사냥한다^[8]. 시각장애인을 대상으로 하는 시각 정보 전달 연구들은 카메라로 촬영한 이미지를 인식 하여 TTS엔진으로 주변 상황을 목소리로 읽어주는 네비게이션 형태, 피부에 전기 자극 포인트를 매트릭스 형태로 장착하여 시각 정보를 전기 자극으로 변화 시켜 전달하는 방법, 뇌에 직접적인 장치를 삽입하여 정보를 전달하는 방법이 연구되고 있다^[9]. 시각장애인 중 특별한 경우 박쥐, 돌고래와 같이 echo를 이용하여 물체를 인식하고 보행을 하는 echolocator가 있다^[10]. 이들은 혀로 ‘딱’, ‘딱’ 하는 클릭 소리를 발생시켜 전방의 물체를 감지하여 거리 및 형태를 인식한다. 다만 이들의 능력을 타인에게 교육을 통해 전달하려는 시도가 있었지만 이러한 후천적 능력을 전달함에 한계가 있었다.

대부분의 이미지 기반 객체 인식은 빛을 이용해서 공간을 감지하는 카메라, LiDAR의 경우 각각의 위치에 빛을 방출하여 반사된 빛을 수신하여 물체를 인식 하는데 좋은 품질의 이미지를 얻기 힘든 환경에서는 예측결과 오류 증가의 원인이 된다. 빛을 측정하는 센서류와 달리 echo 신호를 이미지 정보와 함께 활용하는 경우 빛과 관련된 변화에 따른 영향이 적게 발생하기 때문에 더욱 강건한 출력 결과를 얻는 모델을 생성 할 수 있을 것으로 예상된다. 다양한 chirp신호를 공간에 방출할 경우 공간상의 물체 특성 정보를 반영한 echo를 수신할 수 있을 것이라고 예상할 수 있으며 2채널의 왼쪽/오른쪽으로 수신된 echo 신호를 스펙트럼 이미지로 변환하고 주변 노이즈에 제거에 대한 전처리 작업을 진행하였다. 관측 공간을 Grid의 특정 영역에 물체의 존재 유무를 확률로 출력할 수 있도록 echo 데이터셋을 수집을 진행했으면 이 데이터셋을 통해서 2채널 스펙트럼 이미지에 대해서 예측을 수행하였다. 이를 통해 학습된 모델은 관측 공간 Grid를 가로/세로/깊이 40cm 공간으로 분할했을 때 83%의 정확도를 가진 것으로 결과를 얻을 수 있었다.

논문의 구성의 다음과 같다. 2절에서 indoor 환경의 객체 인식과 관련된 주요 연구를 요약해서 살펴보고 3절에서 제안한 Deep CNN Based Object Detection Using Chirps Emission and Echo Signals in Indoor Environment 에 대해 설명한다. 4절에서는 실험을 위한 데이터셋 수집 공간 설계, echo 데이터수집,

실험결과를 표시하고 마지막 5절 결론에서는 본 연구의 시사점 및 향후 연구 계획을 설명 한다.

II. 관련연구

Indoor 환경에서 객체 인식과 분류, 배경이미지와 분리하는 컴퓨터 비전과 로봇 자율주행 분야에서 중요한 연구 내용이다. Indoor 환경에서 객체 인식과 분류작업은 복잡한 배경 이미지와 배경과 객체 사이의 물리적 거리가 실외보다 짧기 때문에 이미지를 통한 구분을 좀 더 어렵게 한다. 또한 실내 환경의 다양한 변수들에 의해서 영향을 받는데 예를 들면 조명의 변화, 배경 이미지의 기하학적 복잡성의 변화, 배경 이미지 패턴의 복잡성의 변화와 분류 대상 객체 클래스 다양성 등이 있다.

객체와 배경을 분리하기 위해서 사용된 방법으로는 카메라를 통해서 획득한 이미지를 이용하여 전처리 과정과 이미지 분류기법을 이용한 연구가 있으며, 스테레오 카메라, RGB-D 카메라, 초음파 센서, 레이더, 라이다, 음향 센서 등 다양한 센서를 사용하여 획득한 정보를 통합 사용하여 객체 인식 및 분류정확도를 개선하는 연구 방법 등이 있다.

특히 실내 자율주행 로봇에 필요로 하는 Indoor 환경에서의 위치 파악과 객체의 인식, 충돌 회피 경로 계산에 다양한 센서 정보를 활용하는 센서 퓨전이 기법이 주로 사용되고 있다. 이러한 다양한 형태와 대량의 센서 입력 데이터의 분석과 처리를 위해서 인공지능 딥러닝 알고리즘을 활용한 이미지 기반 학습 모델로 객체의 인식과 분류 등에 활용되고 있다.

1.2D Image 이용한 접근법

카메라를 통한 공간 및 객체 인식은 조명변화에 결과 값이 영향을 크게 받으면 배경과 객체를 구분하기가 어렵고, 이를 구분하기 위해서는 높은 계산 비용이 필요하다. 카메라를 이용한 객체 인식은 자율주행 차량 및 Indoor 로봇 주행, 안면인식, 객체 분류, 객체 Segmentation 등에 활용된다. SLAM은 indoor 주행에서 위치 파악을 위해서 사용되는 기술이며^[11], 비지도 학습기반 차원 축소 기법으로는 local deep features alignment (LDFA) 사용된다^[12]. 2D 카메라 2대를 이용하는 스테레오 카메라를 이용하여 이미지 차이를 기반으로 물체의 깊이를 인식하는 기술이 있으며 이와 유사하게 color 이미지와

depth 이미지를 측정하는 키넥트(kinect), RGB-D와 같은 장치를 이용하기도 한다.

2. Point Cloud 기반 접근법

Light detection and ranging (LiDAR) 센서는 자율주행 자동차 및 Automated Guided Vehicle (AGV) 등에서 필요한 고해상도의 3차원 공간 정보 획득 및 지도 생성 및 물체 인식을 위한 중요 센서로 폭넓게 사용되고 있다.

LiDAR 센서를 이용한 데이터를 사용하는 이유로는 적은 계산량으로 비교적 쉽게 3D 공간 정보와 물체를 인식할 수 있는 장점이 있다. 그러나 LiDAR 센서는 고가의 센서이며 레이저 광원의 특성에 따라 안개와 같은 에어로졸이 있는 경우 에어로졸이 렌즈와 같은 역할을 하여 난반사 발생되며 이로 인해 측정 결과 정확도가 떨어지는 문제를 가지고 있다. 최근에는 송신 레이저의 에너지를 증가시켜 이런 난반사에 대응하는 라이다 센서를 통해 항만, 수로 등에서 충돌방지 센서로 활용되고 있다. 수집된 포인트 클라우드를 기반으로 2D 객체 감지 및 3D 감지에 PointNet, PointNET+, Frustum PointNets (F-PointNet)^[13] 방법을 사용가능하며 특히 정렬되지 않은 point를 입력으로 사용 가능하다. 2D 이미지와 포인트 클라우드를 이용한 3D Scene 생성에는 높은 계산 비용이 발생한다.

3. Deep CNN을 기반 접근법

객체인식 분야에서 딥러닝 알고리즘은 인공지능 분야에서 주요 기술로 본 논문에서는 딥러닝 기반 CNN 알고리즘을 이용하여 echo 신호의 주파수 특성과 형태를 학습한다. 논문에서는 indoor 로봇 네비게이션을 위해서 필요한 알고리즘을 설명하고 있다.

Multi-Channel 3D Object Detection CNN는[13] 카메라 영상에서 2D로 프로젝션된 개체로부터 3D 정보를 획득하기 위해서 깊이 정보와 카메라 촬영 이미지와 BEV 이미지 데이터 퓨전을 통해 개선된 3D 객체의 인식 결과를 얻었다. Multi-Channel 3D Object Detection CNN의 경우 3채널의 RGB, 깊이, BEV 이미지를 사용한다. 그러나 이 경우 이미지 카메라 수를 증가 시키면 처리해야 할 데이터의 양 또한 증가하며 센서 비용 및 데이터 처리에 따른 지연과 하드웨어 비용을 증가 시키게 되는 단점이 있다.

III. DECODE: DEep CNN Based Object DEtection Using Chirps Emission and Echo Signals

1. 시스템 아키텍처

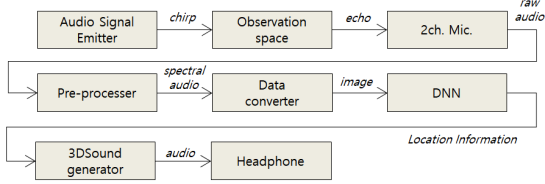


그림 1. DECODE(DEep CNN-based Object DEtection) 시스템의 데이터 Workflow diagram

Fig. 1. Data Workflow diagram of DECODE(DEep CNN-based Object DEtection) system

그림1의 시스템 구성은 관측데이터의 전체적 처리 과정을 나타낸다. 관측자 위치 전방의 임의의 3차원 공간을 관측 공간으로 설정하고 다양한 형태의 chirp신호를 방출한다. chirp신호는 관측공간과 관측자의 거리와 소리의 전달 속도를 감안하여 설정하였으며 chirp 신호는 가청 주파수 범위로 설정하였다. 관측 공간에 방출된 chirp신호에 의해서 echo 신호가 발생되며 발생한 echo를 인간의 상체 모양의 모형 양쪽 귀 위치에 설치한 무지향성 마이크를 통해서 수신한다. 수신한 echo 신호는 잡음 제거와 전처리 작업을 통해서 STFT 변환을 수행하여 2장의 스펙트럼 이미지를 생성하게 되며 생성된 이미지를 DCNN 모델의 입력으로 사용하게 된다. DCNN 모델은 학습된 모델에 의해 결과 값으로 위치정보를 출력하고 이 출력 값으로 3D 사운드 공간의 동일한 위치에 오디오 신호를 헤드폰을 통해서 출력 한다.

2. Chrips 방출과 2채널 echo

Chirp신호는 가청 주파수(20Hz~20KHz) 범위 내에서 반복적인 패턴을 가지는 신호를 선정해야 했는데 이는 인공지능 학습 모델이 정형화된 echo 신호에서 발생하는 차이 값을 인식해서 물체의 특징을 추출하기 때문이다. Chirp 신호는 크기는 echolocator가 발생시키는 click 소리 수준을 감안하여 65db로 설정 하였으며 더미의 턱 아래 위치에서 관측 공간을 향해서 방출 하도록 구성하였다. echo 수신에 사용된 무지향 마이크는 더미의 양쪽 귀에 설치하였으며 양쪽 귀의 물리적인 거리에 따른 소리 도달시간의 차이 값 time difference of arrival (TDOA)과 더미 상체와 머리 형태에 따른 주파

수의 변화 값(머리전달 함수 값) 반영될 수 있도록 하였다. 실험 환경에서는 이전에 방출한 echo사 사라질 수 있도록 chirp 방출 사이에 시간 간격을 1.5초를 적용하였다.

3. Echo signal의 스펙트럼 이미지 변환

모델의 예측 결과 값의 정확성을 향상시키기 위해서는 노이즈를 최소화 하는 방안이 필요하다. 그러나 무향실이 아닌 일반적인 실내 환경에서 노이즈를 제거하기는 쉽지 않으며 벽면, 천장, 바닥 등의 평면에 의해서 반사되는 echo 소리 또한 echo 스펙트럼 이미지 기반 인공지능 학습모델의 예측 정확도에 부정적인 영향을 준다.

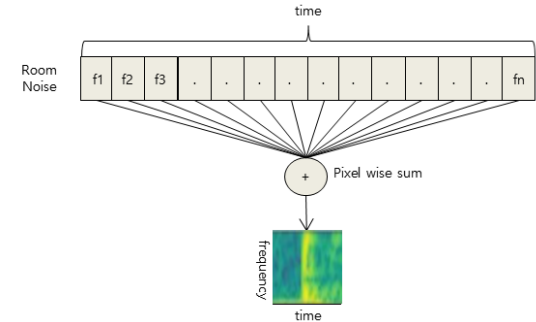


그림 2. 노이즈 제거용 Heatmap 이미지 생성
Fig. 2. Generate heatmap image for noise removal

수신된 echo 신호에 포함된 기본적인 노이즈를 제거 하기 위해서는 방출 신호가 없을때 발생하는 Room 신호를 수집하여 여러 프레임으로 분할하여 주파수 영역에서의 특정 임계값 이상의 에너지를 가지는 부분을 제거 하는 방법을 사용하였다. 이를 위해서 수집된 노이즈 신호를 중첩시켜서 주파수별 에너지 heatmap을 생성하며 생성된 heatmap 이미지를 chirp에 의해서 수신된 echo 스펙트럼 이미지에서 주파수별 마이너스 가중치로 적용하여 노이즈가 echo 스펙트럼 이미지에 적게 반영될 수 있도록 하였다.

4. 딥러닝 CNN 모델 이용한 객체인식

입력으로 사용되는 이미지는 Room 잡음과 chirp 신호와 echo신호가 모두 포함하고 있어서CNN 학습모델이 chirp와 echo사이의 시간차이를 이미지로 학습할 수 있도록 하였다.

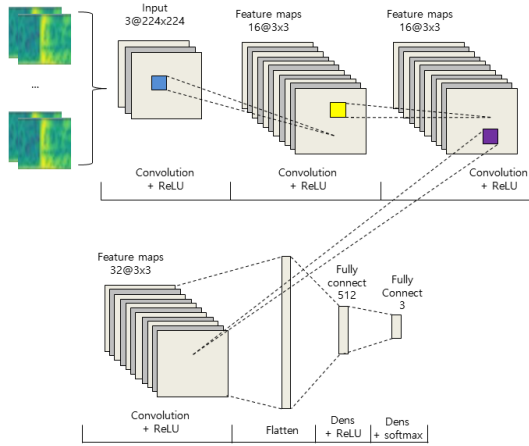


그림 3. Echo 이미지 딥러닝 학습 모델 구조
 Fig. 3. Echo image deep learning model structure

CNN 모델로 3개의 hidden 레이어를 사용했으며 입력 이미지는 왼쪽/오른쪽 3채널 이미지를 한장의 224*224 이미지로 변환하여 모델 입력으로 사용하였다. 오류 함수는 ReLU를 사용하였으면 모델 출력에 Flatten 적용 이후 Dens 512+ReLU, Dens 3 + Softmax를 적용하여 객체의 위치정보를 출력하였다.

IV. 실험 및 결과

1. 실험 공간 설계

실험에 사용된 관측 공간은 2020년 통계청 기준 성인 남자의 평균 눈높이 160cm를 기준으로 하여 관측자의 지상으로부터 높이로 설정 하였고 인간이 물체의 색상을 인식 및 구분할 수 있는 최대 화각인 +60~-60도를 기준으로 보행하는 상황을 고려하여 전방 1미터의 임의의 관측 공간으로 설정 하였다.

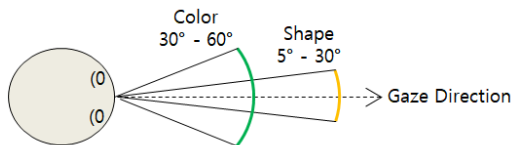


그림 4. 시야각에 따른 식별 대상
 Fig. 4. Identification by viewing angle

그림5와 같이 관측 공간의 크기는 가로 3.4M, 세로 2M, 깊이 3.45M 로 설정하였다.

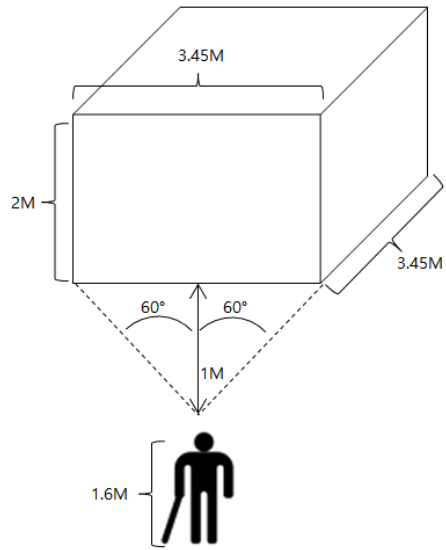


그림 5. 관측 공간
 Fig. 5. Observation space

2. Echo 수집

인공지능 학습용으로 사용한 ehco 데이터를 수집하기 하여 관측공간이 들어갈 수 있는 실내 공간을 선정하고 echo 수집 시 room 노이즈를 제외한 기타 노이즈를 제거하기 위해서 야간에 녹음을 진행하였다.

주요 room 노이즈는 공조기 및 공기정화기 전자제품에서 발생하는 소리가 관측되었으며 비교적 반복적이고 일정한 크기의 잡음 신호임을 확인할 수 있었다.

표 1. 실험에 사용한 chirp 주파수 유형

Table 1. Type of chirp frequency used in the experiment

파형의 형태	Up/Down	시작 주파수	종료주파수
sin	UP	1.5KHz	15KHz
sawtooth	UP	1.5KHz	15KHz
square	UP	1.5KHz	15KHz
sin	DOWN	15KHz	1.5KHz
sawtooth	DOWN	15KHz	1.5KHz
square	DOWN	15KHz	1.5KHz

표1 에서와 같이 echo 수집에 사용된 chirp 신호는 3가지 유형의 주파수 상승/하강하는 형태이며 0.3초 동안 방출 하였다. 관측공간의 물체는 40cm 정사각형 상자를 관측 공간 바닥에 표시는 좌표에 맞추어 이동하면서 1~5층을 쌓으면서 측정하였다. 관측 공간을 기준으로 가로 9번, 세로9번, 높이 5번에 대해서 6가지 chirp신호

에 대해서 10회 chirp 신호를 방출하여 총 48,600장의 이미지와 약 78GB 용량의 데이터를 수집 하였다. 데이터는 파일명을 통해서 구분 가능 하도록 좌표 값과 chirp 유형, 시작 주파수, 종료 주파수, 인덱스, 좌/우 구분의 형태로 저장하여 잡음제거 전처리 프로그램에서 배치 형태로 읽어 들일 수 있도록 하였다.



그림 6. Echo 데이터 수집 환경
Fig. 6. Echo data collection environment

그림6 에서 chirp신호 방출전의 표시된 1.5KHz 미만의 측정값은 기본적으로 수집되는 room 노이즈로 전처리 단계에서 삭제 처리하며 chirp 신호 이후의 값은 잡음제거 알고리즘을 적용 한다.

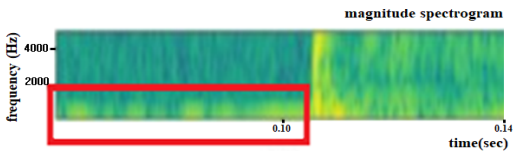


그림 7. 1.5KHz 이하의 잡음신호
Fig. 7. Noise signal less than 1.5KHz

3. 객체 인식 결과

수집된 echo 정보로 DCNN모델의 학습을 진행하였으며 주어진 데이터셋을 기준으로 검증을 총 50 epochs 진행한 결과 약 83% 정도의 정확도를 가지는 것으로 확인된다. 이는 echo 관측 공간이 노이즈가 존재하는 환경임을 감안 하면 chirp 방출에 의해 발생된 echo 신호를 통해서 물체의 인식이 가능하다는 것을 시사한다.

본 논문에서는 비교적 단순한 DCNN 형태의 알고리즘을 사용하였다. 좀 더 깊은 신경망 또는 CNN보다 높은 정확도를 가지는 학습모델 3DCNN, VGG, GoogleNet, LSTM, YOLO 등을 적용하면 더 높은 정확도를 가지는

결과를 얻을 수 있을 것으로 생각된다^[15-16].

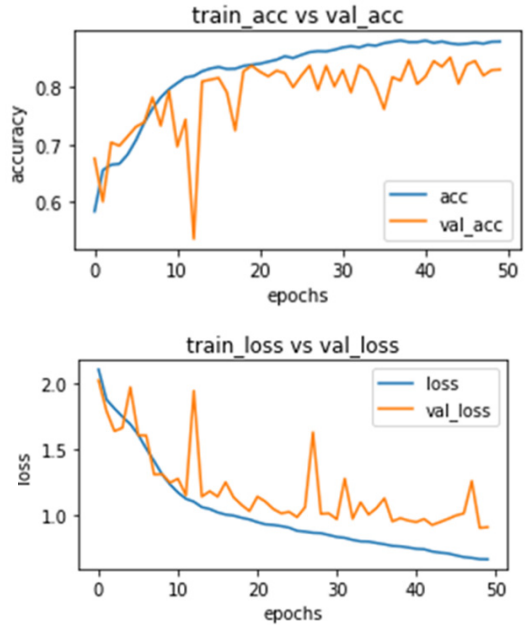


그림 8. echo 이미지 학습모델의 결과 정확성
Fig. 8. Results accuracy of echo image learning model

4. 관측공간과 mapping 된 출력

상기 모델이 출력한 TEXT 형태의 결과 값을 이용하여 Web Audio API의 3D Panning을 이용하여 3D 사운드를 생성하고 동시에 3D UI 공간상의 물체의 위치를 표시하여 관측 공간의 물체의 위치와 3D 소리 출력 값과의 차이를 시각적으로 확인 할 수 있도록 하였다^[17]. UI 구성은 p5.js를 이용하여 구성하였으며 3D 사운드는 p5.js에 있는 Web Audio API 라이브러리인 Sound.js를 사용하였다.

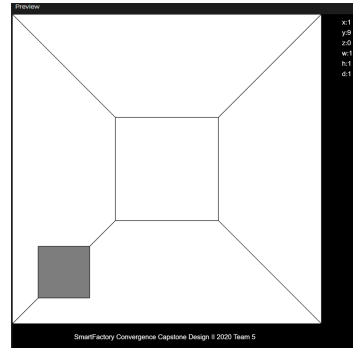


그림 9. 3D 공간의 물체 표시 및 3D 사운드 출력 결과
Fig. 9. Display of objects in 3D space and output of 3D sound

V. 결 론

본 논문에서는 다양한 chirp 신호를 관측대상 공간에 방출하여 발생 되는 echo를 2채널로 수신하여 전처리 과정을 거쳐서 2장의 스펙트럼 이미지를 생성하여 딥러닝 CNN 모델을 통하여 물체의 위치와 형태를 감지하는 방법을 제시 하였다. 또한 모델의 출력 값을 3D 사운드로 출력하여 관측공간의 물체 정보를 3D 사운드 공간과 mapping된 형태로 출력하여 소리를 통해서 이미지 정보를 전달할 수 있는지에 대해 살펴보았다^[16].

이로 인해 echolocator의 물체 감지 능력을 3D 사운드의 형태로 전달 가능한 것을 알 수 있었다.

References

- [1] R. A. Jarvis, "A Perspective on Range Finding Techniques for Computer Vision," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-5, no. 2, pp. 122-139, March 1983
DOI: <https://doi.org/10.1109/TPAMI.1983.4767365>
- [2] Guo, Y.; He, D.; Chai, Li, "A Machine Vision-Based Method for Monitoring Scene-Interactive Behaviors of Dairy Calf MDPI Animals 2020, 10, 190.
DOI: <https://doi.org/10.3390/ani10020190>
- [3] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.
DOI: <https://doi.org/10.1109/CVPR.2016.91>
- [4] D. M. Gavrila and V. Philomin, "Real-time object detection for "smart" vehicles," Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 87-93 vol.1.
DOI: <https://doi.org/10.1109/ICCV.1999.79120>
- [5] Cho, Youngjoon, Kim, Jongwon, "A Study on The Classification of Target-objects with The Deep-learning Model in The Vision-images", Journal of the Korea Academia-Industrial cooperation Society, v.22, no.2, 20-25, February 2021
DOI: <https://doi.org/10.5762/KAIS.2021.22.2.20>
- [6] J. N. Kutz, X. Fu, S. L. Brunton and N. B. Erichson, "Multi-resolution Dynamic Mode Decomposition for Foreground/Background Separation and Object Tracking," 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, 2015, pp. 921-929.
DOI: <https://doi.org/10.1109/ICCVW.2015.122>
- [7] Nenad GUCUNSKI, Zhe WANG, Tong FANG, Ali MAHER, "Rapid Bridge Deck Condition Assessment Using Three Dimensional Visualization of Impact Echo Data", NDTCE'09, Non-Destructive Testing in Civil Engineering Nantes, France, June 30th - July 3rd, 2009.
DOI: [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000060](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000060)
- [8] J. H. Christensen, S. Hornauer and S. X. Yu, "BatVision: Learning to See 3D Spatial Layout with Two Ears," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 1581-1587.
DOI: <https://doi.org/10.1109/ICRA40945.2020.919693>
- [9] M. M. MOORE JACKSON, S. G. MASON, G. E. BIRCH, "Analyzing Trends in Brain Interface Technology: A Method to Compare Studies", Annals of Biomedical Engineering, Vol. 34, No. 5, May 2006 (C 2006) pp. 859-878.
DOI: <https://doi.org/10.1007/s10439-005-9055-7>
- [10] Santani Teng, Amrita Puri, David Whitney, "Ultrafine spatial acuity of blind expert human echolocators", Exp Brain Res(2012) 216:483-488.
DOI: <https://doi.org/10.1007/s00221-011-2951-1>
- [11] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," in IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262, Oct. 2017.
DOI: <https://doi.org/10.1109/TRO.2017.2705103>
- [12] J. Zhang, J. Yu and D. Tao, "Local Deep-Feature Alignment for Unsupervised Dimension Reduction," in IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2420-2432, May 2018.
DOI: <https://doi.org/10.1109/TIP.2018.280421>
- [13] Y. Lian, T. Feng and J. Zhou, "A Dense Pointnet++ Architecture for 3D Point Cloud Semantic Segmentation," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 5061-5064.
DOI: <https://doi.org/10.1109/IGARSS.2019.8898177>
- [14] Wang, L.; Li, R.; Shi, H.; Sun, J.; Zhao, L.; Seah, H.S.; Quah, C.K.; Tandianus, B. "Multi-Channel Convolutional Neural Network Based 3D Object Detection for Indoor Robot Environmental Perception." Sensors 2019, 19, 893.
DOI: <https://doi.org/10.3390/s19040893>
- [15] M. F. Haque, H. Lim and D. Kang, "Object Detection Based on VGG with ResNet Network," 2019 International Conference on Electronics, Information, and Communication (ICEIC), Auckland, New Zealand, 2019, pp. 1-3.
DOI: <https://doi.org/10.23919/ELINFOCOM.2019.8706476>
- [16] Jinsoo Kim, Jeongho Cho. "YOLO-based Real-Time Object Detection Scheme Combining RGB Image with LiDAR Point Cloud," The Journal of Korean Institute of Information Technology, 17(8), 93-105.
DOI: <https://doi.org/10.14801/ikiit.2019.17.8.93>
- [17] A. Mandal, C. V. Lopes, T. Givargis, A. Haghghat, R. Jurdak and P. Baldi, "Beep: 3D indoor positioning

using audible sound," Second IEEE Consumer Communications and Networking Conference, 2005. CCNC. 2005, Las Vegas, NV, 2005, pp. 348-353. DOI: <https://doi.org/10.1109/CCNC.2005.1405195>

저자 소개

Hyunsoo Nam(준회원)



- As of 2021, he is working at YS System Co., Ltd. and is studying for a master's degree in Smart Factory Convergence Department at Sungkyunkwan University. He is interested in Computer Vision and Object detection, Mobile Robot, AIoT
- email: hyunsoo73@g.skku.edu

Jongpil Jeong(정회원)



- 2016~ present: Professor, Dept. of Smart Factory Convergence, Sungkyunkwan University
- 2008: Ph. D. in Computer Engineering (Mobile Computing), Sungkyunkwan University
- 2012~ present: Professor, College of Information and Communication Engineering, Sungkyunkwan University
- 2017~ present: PI, Smart Factory UNIC Center, Sungkyunkwan University
- Interests: Industrial AI, Anomaly detection, Manufacturing data analytics, AI-enabled fault diagnosis and prognosis, 5G-based smart manufacturing, Edge Intelligence
- email: jpjeong@skku.edu