오디오와 레이더를 결합한 딥러닝 환경 분류 연구

김태호, 장준혁*
한양대학교 전자컴퓨터통신공학과
e-mail : kth.address@gmail.com , *jchang@hanyang.ac.kr

A Study on Using Deep learning for Event Classification Based on Audio and Radar

Tae-Ho Kim, Joon-Hyuk Chang*
Department of Electronics and Computer Engineering, Hanyang Univ.

요 약

본 논문에서는 오디오와 레이더 기반의 딥러닝을 활용한 환경 분류 기술을 제안한다. 제안된 환경 분류 기술은 오디오를 이용한 환경 분류 딥러닝 모델과 레이더를 이용한 딥러닝 모델을 앙상블로 결합하여 환경을 분류한다. 특히, 오디오와 레이더 각 성능을 높이기 위해 별도의 모델이 제안된 딥러닝 환경분류 기법은 실내 환경 5 가지를 분류 하였으며, 오디오 또는 레이더 단일 데이터를 활용한 환경분류에 비해 우수한 성능을 보였다.

1. 서론

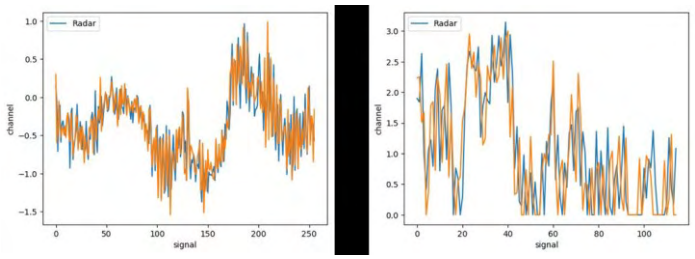
환경 분류는 환경 인지의 한 종류로, 입력 신호에 따라 오디오, 이미지, 동영상 등 다양한 신호를 이용한 분류 기술이 발표되었다. 환경 분류 기술은 각 신호의 특성을 고려한 특징 추출 모델을 통해 신호의 특징 벡터를 추출하고, 이를 이용해 환경의 특성을 모델링하여 어떤 환경인지 분류한다.

딥러닝 모델은 특징벡터 추출 또는 환경 특성 모델링에 사용되며, 다양한 환경 분류 분야에서 우수한 성능을 보였다. 하지만 단일 신호를 통한 환경 분류는 음영지역 발생 및 다른 환경의 유사 신호 발생 문제가 있으며, 이는 분류 성능 저하의 원인이 된다. 본 논문에서는 유사신호 및 음영지역 발생 문제를 해결하기 위해 오디오 신호와 레이더 신호를 결합하여 실내 환경을 분류 하였다. 제안된 환경 분류 기술은 단일 신호 환경 분류 대비 우수한 환경 분류 성능을 보였다.

2. 본론

오디오와 레이더 모델을 결합하기 위해서는 데이터를 동시에 같은 환경에서 수집해야 한다. 이를 위해 실내 주거환경에서 데이터를 수집하였다. 본 연구의 UWB 레이더 데이터 수집은 Novelda 사의 NVA6201-X2 칩을 포함한 NVA-R661 Novelda 개발키트를 활용하였다. 레이더 데이터는 수직 40°, 수평 35°, 거리 3m 내의 환경 변화를 거리 비례로 측정하여 256 단계로 수집하였다. 동시에 오디오 입력을

오디오를 이용한 환경 분류 기술에서는 멜 주파수 켑스트럴 계수(MFCC, Mel Frequency Cepstral Coefficient)를 통해 특징 벡터를 추출한다. 멜 주파수 켑스트럴 계수는 음향 신호의 특징을 잘 표현하여 음향 분류 분야에서 널리 쓰인다[1]. MFCC 추출한 특징 벡터의 인자는 연속적 신호를 특정 시점에서 추출한 것이므로, 시계열 데이터에 적합한 딥러닝 모델을 통해 분류 모델을 설계할 필요가 있다. LSTM(Long-Short Time Memory) 모델은 특징 벡터의 연속적 신호특성을 더 잘 반영할 수 있도록 설계되었다. 따라서 본 논문에서는 오디오 신호의 연속적 특성을 더 잘 반영하기 위해 LSTM 을 활용하여 오디오 분류 모델을 구축하였다.

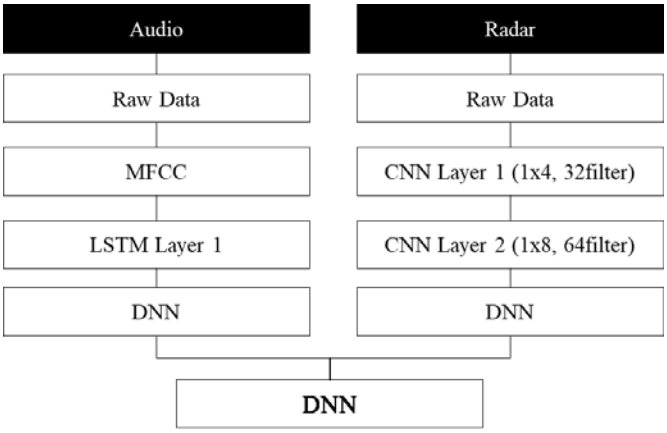


<표 1> Original signal(left) and filtered signal(right)

레이더 데이터는 UWB(Ultra Wideband) 레이더를 통해 수집하였다. 레이더 데이터의 경우 목적에 따라 적합한 필터 및 알고리즘을 사용할 수 있다. 딥러닝을 활용한 레이더 특징추출에서는 최근 CNN (Convolution Neural Network)을 통한 특징추출이 높은 성능을 보였다[2]. CNN 은 이미지 패턴 분석에서 많이 활용된다. 레이더 데이터는 프레임 마다 환경 정보 통한 분석이 가능하다는 점에서 이미지 데이터와 유사하다. 따라서 본 논문에서는 <표 1>과 같이 CNN 을

통해 레이더 데이터의 특징을 추출하고 DNN 과 결합하여 레이더 환경 분류 모델을 구성하였다. 그러나 이미지와 다르게 1 차원 데이터이므로 1 차원 필터를 사용했다.

오디오와 레이더 모델은 각각 시계열 모델과 단일 분류 모델이므로, 모델 결합 시 단일 신호 모델의 결과값에 대한 동기화가 필요하다. 본 논문에서는 시간을 기준으로 동기화하여 결합 모델을 설계하였다.



<표 2> 오디오, 레이더 딥러닝 결합 모델

이는 <표 2>와 같이 나타낼 수 있으며, 최종 결합 모델은 DNN 앙상블로 설명할 수 있다[3]. 이를 통해 단일 신호에서 발생하는 음영지역과 유사 신호 구간에서의 상호 보완적인 학습모델을 설계가 가능하였고 단일 신호 분류 모델 대비 성능 향상을 도모하였다.

오디오 딥러닝 모델과 레이더 딥러닝 모델은 각각 튜닝을 통해 학습 파라미터를 설정하였고 이를 결합하여 최종 분류 모델을 구축하였다. 이를 통해 각 데이터에 최적화된 딥러닝 모델을 구축하였다.

3. 실험결과 및 분석

본 논문에서 제안하는 오디오와 레이더를 결합한 딥러닝 환경 분류 성능 평가를 위하여 수집한 데이터를 학습데이터와 테스트 데이터로 분류하여 진행하였다.

오디오 단일 환경분류의 경우 91.51%의 분류 성능을 보였으며, 레디어 단일 환경분류의 경우 89.25%의 환경 분류 성능을 보였다.

오디오와 레이더 모델을 결합한 환경 분류 모델은 96.47%의 성능을 보여 오디오 환경분류 대비 2.84%, 레이더 환경 분류 대비 5.22%의 성능향상을 보였다.

모델 종류	Accuracy
오디오 분류 모델	91.51%
레이더 분류 모델	89.25%
결합 분류 모델	94.47%

<표 3> 모델 별 성능 비교

특히 레이더 대비 성능 향상이 두드러졌다. 레이더 신호는 다른 환경에서 거리와 부피가 동일한 경우 유사한 신호가 발생하고, 이 경우 분류의 정확도가 낮아 진다. 결합 모델에서는 오디오 신호를 통해 나오는 정보가 추가되어 레이더 유사신호 분류의 정확도가 높아진 것으로 추정된다.

오디오 신호의 경우 잡음과 음향 크기 변화로 인한 왜곡, 공백과 같은 유사 신호 구간의 발생으로 성능 저하가 발생한다. 이 경우 오디오 데이터의 왜곡과 유사 신호 구간에서 레이더 신호가 추가되어 성능이 향상된 것으로 보인다.

4. 결론

본 논문에서는 오디오 신호를 이용한 딥러닝 모델과 레이더 신호를 이용한 딥러닝 모델을 결합한 환경 분류 기술을 도입하였다. 제안하는 환경 분류 기법에서 오디오는 전통적 신호 처리 기법인 MFCC 를 활용하였고, 레이더는 CNN 을 통해 특징벡터를 추출하였다. 각 특징벡터를 시계열 모델인 LSTM 과 딥러닝 모델인 DNN 으로 1 차 분류를 진행한 후, 두 모델을 결합하여 결합 분류 모델을 구성하였다. 결합 분류 모델은 각 신호의 왜곡구간과 유사신호 발생구간을 상호 보완할 수 있도록 설계하였다. 제안하는 기법은 실내 환경데이터를 통해 평가되었으며, 단일 신호 기반의 분류 기법보다 우수한 성능을 보였다.

ACKNOWLEDMENT

이 논문은 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2016-0-00564, 사용자의 의도와 맥락을 이해하는 지능형 인터랙션 기술 연구개발)

참고문헌

[1] Ziyou Xiongt, Regunathan Radhakrishnant, Ajay Divakarant and Thomas S. Huang, COMPARING MFCC AND MPEG-7 AUDIO FEATURES FOR FEATURE EXTRACTION, MAXIMUM LIKELIHOOD HMM AND ENTROPIC PRIOR HMM FOR SPORTS AUDIO CLASSIFICATION, 2003

[2] Seo Yul Kim, Hong Gul Han, Student Member, IEEE, Jin Woo Kim, Sanghoon Lee, Senior Member, IEEE, and Tae Wook Kim, Senior Member, IEEE, A Hand Gesture Recognition Sensor Using Reflected Impulses

[3] Sergey Ioffe, Christian Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift