제한된 계산량으로 가정내 음향 상황을 검출하는 사운드 이벤트 검출 시스템 개발

*장달원 이재원 이종설
전자부품연구원
*dalwon@keti.re.kr

Development of Sound Event Detection for Home with Limited Computation Power

Jang, Dalwon Lee, Jaewon Lee, JongSeol
Korea Electronics Technology Institute

요약

이 논문에서는 가정내 음향 상황에 대한 사운드 이벤트 검출을 수행하는 시스템을 개발하는 내용을 담고 있다. 사운드 이벤트 검출 시스템은 마이크로폰 입력에 대해서 입력신호로부터 특징을 추출하고, 특징으로부터 이벤트가 있었는지 아닌지를 분류하는 형태를 가지고 있다. 본 연구에서는 독립형 디바이스가 가정내 위치한 상황을 가정하여 개발을 진행하였다. 가정내에서 일어날 수 있는 음향 상황을 가정하고 데이터셋 녹음을 진행하였다. 데이터셋을 기반으로 특징과 분류기를 개발하였으며, 적은 계산량으로 결과를 출력해야 하는 독립형 디바이스에 활용하기 위해서 특징셋을 간소화하는 과정을 거쳤다. 개발결과는 가정의 거실환경에서 녹음된 소리를 스피커로 출력하여 테스트하였으며, 다양한 음향 상황에 대한 개발이 추가적으로 필요하다.

1. 서론

음향 입력으로부터 어떤 사건이 발생하였는지, 또는 현재의 상황이 어떤지 알아내는 사운드 이벤트 검출 분야는 최근 데이터셋이 많이 공유되면서 관련 연구의 폭이 넓어지고 있다[1,2]. 또한 AI스피커의 보급으로 인해서 활용을 할 여지도 많이 생기고 있어 관련 연구에 대한 관심이 뜨겁다. 가정 내 다양한 돌발상황을 파악한 후 AI스피커와 같은 가정내 비서기기의의 대응을 위해서 사운드 이벤트 검출기술이 필요하기도 하다.

사운드 이벤트 검출에서도 최근 딥러닝을 이용한 결과가 좋은 성능을 보이고, 특징추출과 분류기 모델이 복잡해지는 경향을 보이고 있다[1]. 하지만, 이 연구에서는 클라우드를 활용한 서버에서 사운드 이벤트 검출을 할 것이 아니라, 독립된 디바이스에서 구동할 예정이다. 따라서 기존의 연구에서와 같이 많이 계산량을 기반으로 검출을 진행할 수 없기에, 제한된 계산량으로 검출을 성공할 수 있는 시스템을 개발하였다. 가정내 일어날 수 있는 다양한 음향상황 중에서 3가지 상황을 가정하고 이를 검출하는 시스템을 개발하였고, 향후 추가적인 개발을 통해서 많은 종류의 음향 상황에 대한 검출이 가능한 시스템을 개발할 예정이다.

2. 시스템 개발

시스템 개발을 위해서 먼저 필요 기능에 대한 정리가 선행된다. 최종적으로 총 10종의 음향상황에 대해서 검출을 수행할 예정이지만, 현 시스템에서는 아래의 3종의 음향상황에 대한 검출을 수행한다. 목표로 하는 음향상황이 정해지면 그에 관련된 데이터를 수집하고, 시스템을 설계해야 한다.

- 강아지 짖는 소리
- 넘어지는 충격음
- 비명 소리

3종의 음향상황에 대한 검출을 수행하기 위해서 데이터셋을 수집해야 한다. 시스템은 가정 내에서 거리가 충분히 떨어진 상황을 가정하고 있기 때문에 실제 가정환경과 같은 곳에서 1-2m 위치에 마이크를 두고 소리를 녹음하였다. 검출되는 상황은 끊임없이 소리의 발생여부를 확인해야 하기 때문에 주변잡음에 대한 데이터셋도 구축하였다.

〈녹음 현장〉

넘어지는 충격음 같은 경우에는 굉장히 짧은 시간 (<50ms) 안에 일어나는 소리이기 때문에, 이를 검출해내기 위해서는 짧은 시간 단위 프레임을 사용하여야 한다. 이를 위해서 16kHz 입력에 대해서 512 pt 길이의 프레임을 256pt마다 이동시키는 형태로 구현을 하였다. 짧은

길이의 프레임을 사용하기 때문에 한 프레임에 대한 오류가 있을 수 있기에, 후처리를 통해서 최근 프레임의 결과를 모아서 최종적으로 결정을 내는 과정이 필요하다.

효율적인 시스템 개발을 위해서 1차적으로 PC상에서 많은 계산량에 기반하여 특징과 분류기를 시험하고, 2차적으로 그 중 중요한 특징과 심플한 분류기 형태의 시스템을 구성해서 개별 디바이스에서 구동되게 한다. PC상에서 개발할 때 사용한 특징은 Mel-Frequency Cepstral Coefficients (MFCC), 선형 예측 계수 (Linear prediction coefficients, LPC), 에너지, 주파수 중심값, 제로 크로싱 레이트 (zero-crossing rate), 스펙트랄 콘트라스트(spectral contrast), 총 6가지 32차의 특징을 사용하였다. 성능은 support vector machine (SVM) 분류기와 neural network (NN) 분류기를 사용하여 검증하였다. 각각의 음향 상황에 대해서 잡음과 음향 상황을 1:1로 분류하는 분류기를 3가지 만들었고, 각 분류기는 평행하게 동작한다.

사용한 특징들은 서로 상호보완적이기는 하나, 중복된 정보를 가질 수도 있기에, 모든 것을 다 사용하는 것은 계산량이 제한된 디바이스에는 적합하지 않다. PC상에서 이루어진 연구에서 각 특징별로 분류 성능을 살펴보았고, 독립 디바이스에서는 MFCC, 주파수 중심값, 제로 크로싱 레이트, 세 가지 특징만 사용하였다. 그리고, 분류기도 가장 낮은 계산량을 가지는 선형 커널을 사용하는 SVM을 활용하였다. 선형 커널을 가지는 SVM은 특징의 차수만큼의 곱셈과 덧셈, 그리고 비교 한번으로 분류기를 동작할 수 있다.

3. 테스트

전장에서 밝혔듯이 개발을 위해서 1차적으로 PC상에서 구현되는 모듈을 개발하였고, 이를 경량화해서 개별 디바이스에 적용하는 과정을 거쳤다. 개발을 두 단계로 진행하였기 때문에, 두 가지 형태의 결과를 제시한다. PC상에서의 결과는 잡음 상황과 음향상황이 일어나는 3가지 상황에 대해서 각각의 분류기의 성능을 비교하였다. 개별 디바이스에서는 실험은 가정의 거실과 같은 환경의 테스트룸에서 스피커를 통해서 검출이 되어야 할 소리를 출력하고 2m 가량 떨어진 곳에 검출 디바이스를 두고 성능을 실험하였다.

〈테스트 과정: 우측 - 스피커, 좌측 - 디바이스〉

PC상에서 구현하였을 때의 성능 결과는 아래와 같이 두 가지 분류기에 대해서 모든 특징을 다 사용했을 때의 성능을 제시한다. 음원 녹음된 데이터셋 중 테스트셋을 기준으로 프레임 단위로 성능을 측정하였다.

음향 상황	SVM 분류기			NN 분류기
	linear	다항	RBF	
강아지 짖는 소리	98.7	98.6	96.3	99.5
넘어지는 충격음	99.7	99.6	99.8	99.8
비명 소리	99.2	98.9	99.6	99.3

실제 테스트에서는 성능 실험 결과 아래와 같은 결과가 나왔고, 넘어지는 충격음에 대해서는 성능이 떨어짐을 확인하였다. 모든 경우에 false alarm은 일어나지 않았으며, 검출이 실패한 경우는 다른 음향상황으로 검출하거나 검출을 못 하는 두 가지 상황이 있었으며, 여기에서는 두 가지 상황에 대해서 따로 분리하지는 않았다.

음향 상황	시행	검출 성공
강아지 짖는 소리	10	9
넘어지는 충격음	9	3
비명 소리	8	5

실제 테스트에서는 성능이 떨어진 이유로는 하드웨어에 대한 이해가 부족한 점을 들 수 있다. 개발 디바이스는 작은 크기의 MEMS 마이크로폰은 사용하고, 이것은 데이터셋 개발을 위해 사용하던 마이크로폰에 비해서 성능이 많이 떨어짐이 당연하다. 특히나 해상력이 많이 떨어질 것이고 이런 차이가 넘어지는 충격음 같이 짧은 시간에 강한 강도로 발생하는 음향을 찾아내지 못하는 것으로 판단된다. 추후 연구에서는 개발 디바이스에 적합한 사운드 이벤트 검출 시스템을 개발할 예정이다.

4. 결론

이 논문에서는 가정내 사용하는 계산량이 제한된 디바이스에서의 사운드 이벤트 검출 시스템을 개발하는 과정에 대해 논하였다. 음향상황을 가정하고, 데이터셋을 모으고, 특징셋과 분류기를 연구하여 적합한 특징셋과 분류기를 시스템에서 활용하였다. PC상에서 녹음된 음원을 이용하였을 때는 좋은 성능이 나왔지만, 실제 가정환경에서 테스트하였을 때의 성능은 만족스럽지 않았다. 문제점을 보완하는 추가개발이 필요하다.

감사의 글

본 연구는 산업통상자원부의 "전자시스템전문기술개발사업"의 지원을 받아 수행된 연구결과임 (20000111, 2018)

참고문헌

[1] D. Jang, J. Lee, M. Kim, and J.S. Lee, “A survey paper on category and method of sound signal analysis,” ICONI 2018.

[2] A Mesaros, T Heittola, and T Virtanen, “TUT database for acoustic scene classification and sound event detection,” EUSIPCO 2016