



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0064018  
(43) 공개일자 2021년06월02일

(51) 국제특허분류(Int. Cl.)  
G10H 1/00 (2006.01) G08B 13/16 (2006.01)  
(52) CPC특허분류  
G10H 1/00 (2013.01)  
G06K 9/3233 (2013.01)  
(21) 출원번호 10-2020-0035181  
(22) 출원일자 2020년03월23일  
심사청구일자 2020년03월23일  
(30) 우선권주장  
62/940,058 2019년11월25일 미국(US)

(71) 출원인  
광주과학기술원  
광주광역시 북구 첨단과기로 123 (오룡동)  
(72) 발명자  
김홍국  
광주광역시 북구 첨단과기로 123(오룡동) 광주과학기술원 전기전자컴퓨터공학부  
박인영  
광주광역시 북구 첨단과기로 123(오룡동) 광주과학기술원 전기전자컴퓨터공학부  
(74) 대리인  
김기문

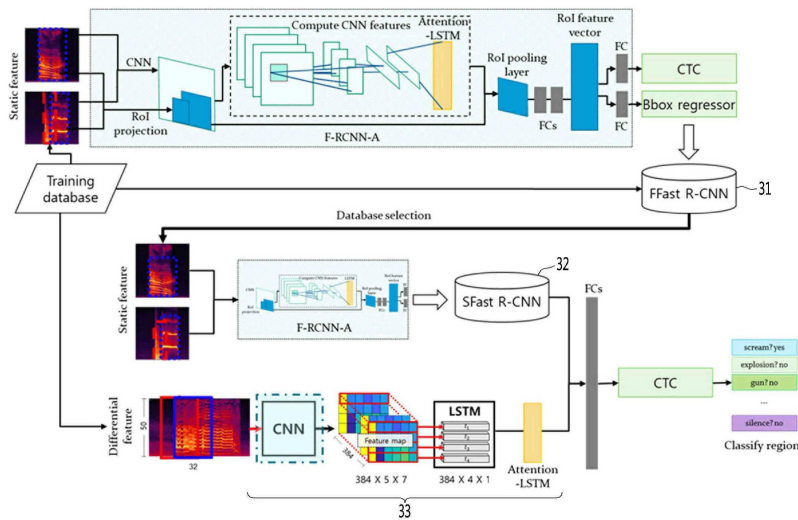
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 딥러닝 기반 감지상황에서의 음향 사건 탐지 방법

(57) 요약

본 개시는 적어도 하나 이상의 음원이 포함된 음향 데이터에 기초하여 사의 음향 데이터에 포함된 복합특징값을 추출하고, 고속 지역 합성곱 장단기 기억 신경망(Fast Region-CNN-LSTM)을 기반으로 하는 인공 지능 모델을 이용하여 상기 복합특징값에 포함된 적어도 하나 이상의 음원을 분류하고, 분류된 음원을 이용하여 사건을 감지하는 방법을 개시한다.

대표도



(52) CPC특허분류

- G06K 9/481 (2013.01)
- G06K 9/627 (2013.01)
- G06N 3/04 (2013.01)
- G06N 3/08 (2013.01)
- G08B 13/1672 (2013.01)
- G10H 2210/031 (2013.01)
- G10H 2250/131 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711097839(NN24670)
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	인공지능SW 선도기술 및 유망 신기술 개발
연구과제명	드론을 활용한 재난 대응을 위한 기계학습 기반 음향지능 기술 개발
기 여 율	1/1
과제수행기관명	광주과학기술원
연구기간	2019.07.01 ~ 2020.03.31

---

## 명세서

### 청구범위

#### 청구항 1

음향 데이터를 이용한 사건 감지 방법에 있어서  
적어도 하나 이상의 음원이 포함된 음향 데이터를 수신하는 단계;  
상기 음향 데이터에 포함된 복합특징값을 추출하는 단계;  
인공 지능 모델을 이용하여 상기 복합특징값에 포함된 적어도 하나 이상의 음원을 각각 분류하는 단계; 및  
상기 분류된 음원 각각의 조합에 기초하여 사건을 감지하는 단계를 포함하는,  
사건 감지 방법.

#### 청구항 2

제1항에 있어서,  
상기 복합특징값은  
수신된 음향 데이터의 스펙트로그램(spectrogram)인 정적 특징(static feature) 및 상기 음향 데이터와 기 설정된 시간 이전의 음향 데이터의 차이에 기초한 차등 특징(differential feature)을 포함하는,  
사건 감지 방법.

#### 청구항 3

제2항에 있어서  
상기 인공 지능 모델은,  
고속 지역 합성곱 장단기 신경망 (Fast R-CNN-LSTM) 을 전대역(full-band height)에서 레이블링되지 않은 학습 데이터를 사용하여 사전 학습시킨 FFast R-CNN-LSTM, 상기 FFast R-CNN-LSTM의 학습 데이터를 선별하여 학습시킨 SFast R-CNN-LSTM 및 차등 특징이 반영되는 인공지능 모델을 포함하는,  
사건 감지 방법.

#### 청구항 4

제 3항에 있어서  
상기 인공 지능 모델을 이용하여 상기 복합특징값에 포함된 음원을 분류하는 단계는,  
상기 FFast R-CNN-LSTM을 이용하여 상기 정적 특징의 콘볼루션 특징맵을 획득하고, 상기 콘볼루션 특징맵으로부터 기 설정된 관심 영역(Region of interest)의 특징 벡터를 추출하여 음원을 분류하는 단계를 포함하는,  
사건 감지 방법.

#### 청구항 5

제 4항에 있어서  
상기 콘볼루션 특징맵으로부터 기 설정된 관심 영역(Region of interest)의 특징 벡터를 추출하여 음원을 분류하는 단계는,  
상기 FFast R-CNN-LSTM이 출력한 기 설정된 관심 영역의 특징 벡터에 기초하여 상기 관심 영역을 재설정하는 단계; 및  
상기 SFast R-CNN-LSTM이 상기 정적 특징의 콘볼루션 특징맵으로부터 재설정된 관심 영역의 특징 벡터를 추출하

여 상기 복합특징값에 포함된 음원을 분류하는 단계를 더 포함하는,  
사건 감지 방법.

#### 청구항 6

제5항에 있어서,

상기 분류된 음원 각각의 조합에 기초하여 사건을 감지하는 단계는,

CTC(Connectionist Temporal Classification)를 이용하여 상기 음향 데이터의 각 시간 단계(time-step) 또는 프레임 별로 레이블링 하여 최종 사건을 감지하는 단계를 포함하는,

사건 감지 방법.

#### 청구항 7

제3항에 있어서

상기 인공 지능 모델을 이용하여 상기 복합특징값에 포함된 음원을 분류하는 단계는,

상기 차등 특징이 반영되는 인공지능 모델이 상기 차등 특징의 특징 맵을 획득하고, 상기 차등 특징의 특징 맵 으로부터 특징 벡터를 추출하고 상기 특징 벡터에 포함된 음원을 분류하는 단계를 포함하는,

사건 감지 방법.

#### 청구항 8

제7항에 있어서,

상기 분류된 음원 각각의 조합에 기초하여 사건을 감지하는 단계는,

CTC(Connectionist Temporal Classification)를 이용하여 상기 음향 데이터의 각 시간 단계(time-step) 또는 프레임 별로 레이블링 하여 최종 사건을 감지하는 단계를 포함하는,

사건 감지 방법.

#### 청구항 9

제3항에 있어서

상기 인공 지능 모델을 이용하여 상기 복합특징값에 포함된 음원을 분류하는 단계는,

상기 FFast R-CNN-LSTM이 상기 정적 특징의 콘볼루션 특징맵을 획득하고, 상기 콘볼루션 특징맵으로부터 기 설정된 관심 영역(Region of interest)의 특징 벡터를 추출하고, 상기 기 설정된 관심 영역을 재설정하는 단계;

상기 SFast R-CNN-LSTM이 상기 정적 특징의 콘볼루션 특징맵을 획득하고, 상기 재설정된 관심 영역의 특징 벡터를 추출하는 단계;

상기 차등 특징이 반영되는 인공지능 모델이 상기 차등 특징의 특징 맵을 획득하고, 상기 차등 특징의 특징 맵 으로부터 특징 벡터를 추출하는 단계 및

상기 특징 벡터에 포함된 음원을 분류하는 단계; 및

CTC(Connectionist Temporal Classification)를 이용하여 상기 음향 데이터의 각 시간 단계(time-step) 또는 프레임 별로 레이블링 하여 최종 사건을 감지하는 단계를 포함하는,

사건 감지 방법.

### 발명의 설명

### 기술 분야

[0001] 본 개시는 딥러닝 기반 감지상황에서의 음향 사건 탐지 방법 및 시스템에 관련된 것으로, 본 개시는 서로 다른 특징을 지닌 음향 신호의 주파수 영역을 탐지하여 고속 지역 합성곱 기반 네트워크(F-R-CNN)를 사용함으로써,

거리와 관계없이 실시간으로 복합적인 음향을 검출할 수 있다.

### 배경 기술

- [0002] 음향 사건을 검출하고 음향 사건의 종류를 분류하는 기술은, 상황-인지(context-aware) 기술과 융합되어 사용자의 주변 환경 판단에 적용되기 위해서 꾸준히 연구되어 왔다.
- [0003] 이러한 연구에 따른 종래의 단일 음향 감지 모델은 확연히 드러나는 음향만 감지하므로 동시에 발생하는 복합 음원에 대하여 실제 감지를 진행할 때 정보가 훼손되는 문제가 있었다. 즉, 복합 음원 감지의 정확도가 낮은 문제가 있었다.
- [0004] 그리고, 종래의 음향 사건 감지 시스템은 환경에 따라 감지해야하는 음원보다 큰 잡음이 검출되는 경우 감지 환경의 정확도가 낮아지는 문제가 있었고, 음원이 발생하는 지점으로부터의 거리에 따라 음원 검출의 정확도가 낮아지는 문제가 있었다.

### 발명의 내용

#### 해결하려는 과제

- [0005] 본 개시의 목적은 고속 지역 합성곱 기반 네트워크(F-R-CNN)을 이용하여 인공 지능 모델을 학습시키고, 감지상황에서의 실시간 음향 사건 감지를 목적으로 한다.
- [0006] 본 개시는 거리에 관계없이 음향 특징을 추출하고 음향감지상황에서 복수개의 음향 발생을 검출하는 복합 음향 사건 감지를 목적으로 한다.

#### 과제의 해결 수단

- [0007] 본 개시는 음향 데이터를 이용한 사건 감지 방법에 있어서 적어도 하나 이상의 음원이 포함된 음향 데이터를 수신하는 단계; 상기 음향 데이터에 포함된 복합음향 특징을 추출하는 단계; 인공 지능 모델을 이용하여 상기 복합음향 특징에 포함된 적어도 하나 이상의 음원을 각각 분류하는 단계; 및 상기 분류된 음원 각각의 조합에 기초하여 사건을 감지하는 단계를 포함하는 사건 감지 방법을 개시한다.
- [0008] 또한, 본 개시에 있어서 상기 인공 지능 모델을 이용하여 상기 복합음향 특징에 포함된 음원을 분류하는 단계는, 상기 정적 특징의 콘볼루션 특징맵을 획득하고, 상기 콘볼루션 특징맵으로부터 기 설정된 관심 영역 (Region of interest)의 특징 벡터를 추출하여 음원을 분류하는 단계 및 상기 차등 특징의 특징 맵을 획득하고, 상기 차등 특징의 특징 맵으로부터 특징 벡터를 추출하고 상기 특징 벡터에 포함된 음원을 분류하는 단계를 포함할 수 있다.

#### 발명의 효과

- [0009] 본 개시는 고속 지역 합성곱 기반 네트워크(F-R-CNN)을 사용하여 영상 정보를 이용한 상황 감지와 더불어 음향 정보에 기초한 개선된 실시간 감지를 제공할 수 있다.
- [0010] 또한, 본 개시는 거리에 관계없이 음향 특징을 파악하므로, 기존의 음향 기반 감지시스템의 단점을 보완하고 복수개의 음향 발생을 검출하여 정확한 음향 사건 발생을 감지할 수 있다.

#### 도면의 간단한 설명

- [0011] 도 1은 본 개시의 일 실시 예에 따른 순서도를 나타낸다.
- 도 2는 본 개시의 일 실시 예에 따른 알고리즘 진행과정을 나타낸다.
- 도 3은 본 개시의 일 실시 예에 따른 알고리즘 진행과정을 나타낸다.
- 도 4는 본 개시의 일 실시 예에 따른 Attention-LSTM을 나타낸다.
- 도 5는 본 개시의 일 실시 예에 따른 CTC(Connectionist Temporal Classification)를 나타낸다.
- 도 6은 본 개시의 테스트 결과에 따른 성능을 나타낸다.
- 도 7은 본 개시의 테스트 결과에 따른 성능을 나타낸다.

도 8은 본 개시의 테스트 결과에 따른 성능을 나타낸다.

**발명을 실시하기 위한 구체적인 내용**

- [0012] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시 예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.
- [0013] 첨부된 도면은 본 명세서에 개시된 실시 예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않으며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0014] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0015] 본 개시에서 사용될 용어의 표현을 정의하면 아래와 같다.
- [0016] 음향 사건 감지(acoustic event detection, AED), 지역 합성곱 신경망(regional-convolutional neural networks, R-CNN), 장단기 기억 신경망(long short-term memory ,LSTM), CTC(connectionist temporal classification).
- [0017] 본 개시는 다양한 잡음 환경 및 거리에서 작동하는 음향 사건 감지 시스템을 제안한다. 그 방법은 잡음환경에서 서로 다른 특징을 지닌 시계열 음향 사건의 주파수 영역을 탐지하는 지역 합성곱 기반 네트워크(R-CNN)를 사용한다. 이 때, 고속 지역 합성곱 기반 네트워크(F-R-CNN)를 사용함으로써 본 개시는 실시간으로 구현될 수 있다.
- [0018] 이를 통해 주파수대역에서의 관심지역 정보를 일일이 수작업으로 라벨링하지 않고도 준지도학습(semi-supervised learning)을 가능하게 한다.
- [0019] 또한, 본 개시에 따르면 사건 감지 성능이 거리에 따른 신호의 세기에 좌우되는 문제를 해결하기 위해 인공 지능 모델의 입력값은 복합특징값(multi-feature)을 사용하도록 설계될 수 있다.
- [0020] 부가적으로 신호의 시계열적 특징을 위해 장단기 기억 신경망이 적용되었으며, 어텐션(Attention) 알고리즘을 통하여 음원의 길이적 특성을 반영하여 음원 감지 성능을 향상시켰다. 최종적으로, CTC(connectionist temporal classification) 를 적용하여 복합 음원이 포함된 하나의 사건을 감지한다.
- [0022] 이하 도 1에 대하여 설명한다.
- [0023] 도 1은 본 개시의 순서도를 나타낸다.
- [0024] 도 1을 참조하면, 본 개시의 음향 사건 감지 시스템은 적어도 하나 이상의 음원이 포함된 음향 데이터를 수신하는 단계(S10)를 포함할 수 있다.
- [0025] 구체적으로 음원이란 음(sound)의 근원(source)을 포함할 수 있으며, 음향 사건 감지 시스템은 적어도 하나 이상의 음원을 포함하는 음향 데이터를 포함할 수 있다.
- [0026] 예를 들어 하나의 음향 데이터에는 대화 음원, 물체가 부딪히는 음원, 총성 등 특정 물체 또는 생물로부터 발생하는 모든 소리 등을 포함될 수 있다. 그리고 상기 음향 데이터는 음원으로부터 추출된 음향 특징 데이터를 포함할 수 있다.
- [0027] 이때 정확한 음향 데이터를 획득하기 위하여 잡음 제거가 전처리 과정에서 수행될 수 있으며, 잡음의 스펙트럼 주파수 대역이 발생한 사건과 다르거나 각 사건(event)에 고유한 주파수 대역이있는 경우 R-CNN (region-convolutional neural network)을 이용할 수 있다.
- [0028] 본 개시의 음향 사건 감지 시스템은 상기 음향 데이터에 포함된 복합특징값을 추출하는 단계(S20)를 포함할 수 있다.
- [0029] 이때, 상기 복합특징값은 수신된 음향 데이터의 스펙트로그램(spectrogram)인 정적 특징(static feature) 및 상기 음향 데이터와 기 설정된 시간 이전의 음향 데이터의 차이에 기초한 차등 특징(differential feature)을 포함할 수 있다.

- [0030] 구체적으로 정적 특징(static feature)는 음향 데이터의 특징이 포함된 로그 멜 밴드(Log Mel-Band) 에너지 기반 이미지를 의미할 수 있다.
- [0031] 이때, Log Mel-Band 에너지는 음향 신호의 특성을 잘 나타낼 수 있는 에너지이다. 그러나, 복합적인 음향 신호를 분석하는데 한계가 있다. 음향 데이터는 일반적으로 데이터의 세기(power)에 의해 정규화되지만 때로는 거리에 따른 세기가 약해지거나 잡음이 섞여있는 경우 특징 정보가 손상될 수 있다.
- [0032] 따라서 본 개시에서는 음향 데이터에 시계열 특성이 있다는 점을 고려하여 현재 음향 데이터에서 이전 데이터와의 차이를 통하여 로그 멜 밴드 에너지의 변화를 측정 할 수 있다.
- [0033] 상기 방법을 통하여 생성된 차등 특징(differential feature)은 현재 음향 데이터와 기 설정된 시간 이전의 음향 데이터가 이미지로 표현된 스펙트럼 간의 변화량에 기초하므로 음성 데이터의 발생 지점으로부터 거리에 관계없이 음향 데이터의 특징을 반영할 수 있다.
- [0034] 복합특징값을 획득한 이후 본 개시는 인공 지능 모델을 이용하여 상기 복합특징값에 포함된 적어도 하나 이상의 음원을 각각 분류할 수 있다(S30). 그리고 분류된 음원 각각의 조합에 기초하여 사건을 감지할 수 있다(S40).
- [0035] 이하 본 개시의 알고리즘과 관련하여 도 2 내지 도 5에서 구체적으로 설명한다.
- [0036] 본 개시는 다양한 유형의 딥 러닝 출력 계층에 대한 지역(region) 정보를 사용하는 Fast R-CNN-LSTM 기반 실시간 감지 시스템을 개시한다.
- [0037] 구체적으로 본 개시에 따르면, 고속 지역 합성곱 장단기 기억 신경망(Fast R-CNN-Attention LSTM)을 음향 데이터가 이미지로 표현된 학습데이터를 이용하여 학습시켜 FFast R-CNN-LSTM(31)모델 생성할 수 있다.
- [0038] 본 개시는 생성된 FFast R-CNN-LSTM(31)모델의 디코딩 값이 일정 기준값 이상의 스코어를 갖는 학습 데이터를 선별하고, 선별된 학습 데이터를 이용하여 준지도 지역 라벨링(semi-supervised region labeling)을 통해 SFast R-CNN-LSTM(32) 모델을 생성할 수 있다. 상기 과정으로 생성된 SFast R-CNN-LSTM(32)모델은 Spectrum 구간에서 음원의 특징을 나타내는 지역(region)을 모델 스스로 추출할 수 있으므로 핸드 라벨링 비용을 최소화할 수 있다.
- [0039] 이하 도 2에서 구체적으로 설명한다.
- [0040] 일반적으로 CNN(Convolution Neural Network)은 이미지 처리에 적합한 인공 지능 모델이다. 본 개시에서는 CNN의 입력으로 음향 데이터를 사용하기 위해 여러 시점을 수집하고 이미지 처리에 사용하기 위해 음향 데이터의 여러 프레임을 하나의 이미지로 구성할 수 있다.
- [0041] 그리고 본 개시의 인공 지능 모델은 CNN과 RNN (LSTM)을 연결하는 구조의 음향에 따른 사건 감지 알고리즘을 포함할 수 있다.
- [0042] 도 2를 참조하면, 본 개시의 알고리즘은 FFast R-CNN-LSTM(31) 모델과 SFast R-CNN-LSTM 모델(32) 및 차등 특징이 반영되는 인공지능 모델(33)을 포함할 수 있다.
- [0043] 이때, FFast R-CNN-LSTM (31)은 기존의 고속 지역 합성곱 신경망(Fast R-CNN-LSTM)을 전대역(full-band height)에서 레이블링되지 않은 학습 데이터를 사용하여 사전 학습시킨 모델일 수 있다.
- [0044] 구체적으로 고속 지역 합성곱 신경망(Fast R-CNN)은 이미지로부터 특징데이터가 존재할 위치에 Bounding Box Proposal이 모두 생성되는 R-CNN의 병목(bottleneck)구조의 단점을 개선하고자 제안된 방식을 의미할 수 있다. 구체적으로 이전의 R-CNN 과 비교하여 각 Bbox-Proposal들이 CNN을 거치는것이 아니라, 전체 이미지에 대해 CNN을 한번 거친 후 출력 된 특징 맵(feature map)단에서 객체 탐지를 수행하는 것을 특징으로 한다.
- [0045] 또한, 본 개시는 SFast R-CNN-LSTM 모델(32)을 포함할 수 있다.
- [0046] 이때, SFast R-CNN-LSTM(32)모델은 Spectrum 구간에서 음원의 특징을 나타내는 지역(region)을 모델 스스로 추출할 수 있다.
- [0047] 구체적으로 본 개시는 SFast R-CNN-LSTM 모델(32)을 훈련시키기 위하여 FFast R-CNN-LSTM 의 테스트 결과 기 설정된 값 이상의 스코어를 획득한 적절한 훈련 데이터를 선별할 수 있다. 또한, 본 개시의 SFast R-CNN-LSTM 모델(32)는 준지도 영역 레이블링(semi-supervised region labeling)을 사용하는 Fast R-CNN-LSTM을 포함할 수 있다. 본 개시의 SFast R-CNN-LSTM 모델(32)은 선별된 훈련 데이터를 사용하므로 핸드 라벨링 비용을 최소화할

수 있다.

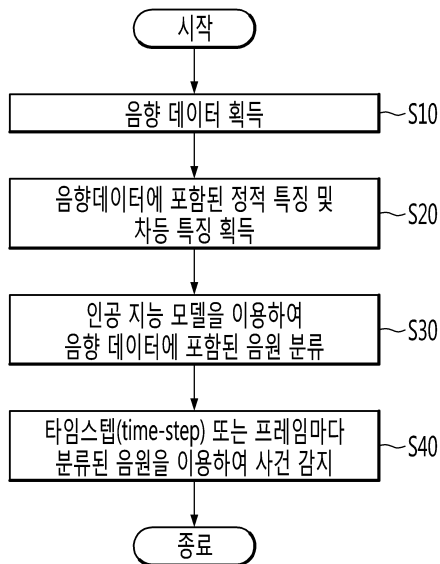
- [0048] 그리고 본 개시의 감지 시스템은 차등 특징이 반영되는 인공지능 모델(33)을 포함할 수 있다.
- [0049] 이때, 차등 특징이 반영되는 인공지능 모델(33)은 SFast R-CNN-LSTM 모델(32)와 병렬적으로 학습될 수 있다. 또한 SFast R-CNN-LSTM 모델(32)과 음향 감지를 위한 일부 층(layer)를 공유하여 하나의 사건 감지를 수행할 수 있다.
- [0050] 이때 일부 층(layer)는 CTC 알고리즘 층을 포함할 수 있다.
- [0051] 상기 차등 특징이 반영되는 인공지능 모델(33)은 CNN과 LSTM이 연결된 층(layer)을 포함하고, 차등 특징이 CNN과 RNN(LSTM)이 연결된 알고리즘에 제공되면 차등 특징이 반영된 특징 벡터를 생성할 수 있다. 이때 차등 특징은 음원에서 추출한 특정시간(T시간)의 음향데이터의 이미지 정보에서 특정 시간 이전(T-1시간) 음향 데이터의 이미지 정보를 뺀 변화량을 포함하고, 차등 특징이 반영되는 인공지능 모델(33)은 상기 차등 특징값을 사용하여 학습될 수 있다.
- [0052] 이하 도 3을 설명한다.
- [0053] 도 3을 참조하면, 본 개시의 음향 감지 시스템의 진행 과정을 나타낸다.
- [0054] 본 개시의 음향 감지 시스템은 적어도 하나 이상의 음원이 포함된 음향 데이터를 수신하고, 상기 음향 데이터에 포함된 복합특징값을 추출하는 할 수 있다(S20). 그리고 추출된 복합특징값 중 정적 특징(static feature)은 FFast R-CNN-LSTM 모델(31)에 입력되고, FFast R-CNN-LSTM 모델(31)의 출력 결과 S20단계에서 설정 되어 있었던 기 설정된 관심 영역(RoI)이 재설정될 수 있다.
- [0055] 본 개시의 감지 시스템은 음향 데이터의 정적 특징을 SFast R-CNN-LSTM 모델(32)에 제공하고, 음향 데이터의 정적 특징의 콘볼루션 특징맵으로부터 재설정된 관심 영역의 특징 벡터를 추출하여 상기 복합특징값에 포함된 음원을 분류할 수 있다.
- [0056] 또한, SFast R-CNN-LSTM 모델(32)과 병렬적으로 차등 특징이 반영되는 인공지능 모델(33)은 음향 데이터의 차등 특징(differential feature)를 이용하여 콘볼루션 특징맵으로부터 특징 벡터를 추출하고, 상기 복합특징값에 포함된 음원을 분류할 수 있다.
- [0057] 이때, 차등 특징이 반영되는 인공지능 모델(33)과 정적 특징이 반영된 SFast R-CNN-LSTM 모델(32)는 특정 층(layer)을 공유하여 상기 복합특징값에 포함된 적어도 하나 이상의 음원을 각각 분류하는 단계(S30) 및 상기 분류된 음원 각각의 조합에 기초하여 사건을 감지하는 단계(S40)를 포함할 수 있다.
- [0058] 구체적으로 공유되는 특정 층(layer)는 시계열적 특징 및 지역적 특징이 반영되는 Attention-LSTM 알고리즘과 CTC(Connectionist Temporal Classification) 알고리즘을 포함할 수 있다.
- [0059] 이때, Attention-LSTM 알고리즘과 CTC 알고리즘은 도 2 에 나타난 모델(31,32,33)각각에 포함되어 음향 감지를 수행할 수 있다.
- [0060] 이하 Attention-LSTM 및 CTC 알고리즘을 도 4내지 도 5에서 설명한다.
- [0061] 도 4는 본 개시의 Attention 알고리즘을 나타낸다.
- [0062] 먼저 본 개시의 사건 감지에서 획득되는 음향 데이터는 다양한 길이의 사건(Event)를 감지할 수 있어야 한다.
- [0063] 본 개시는 음향 데이터의 지역(region)적 특징을 특징맵에서 추출하고 Attention 알고리즘을 적용하여 상기 지역적 특징에 특정 가중치는 부여하여 음향 감지 정확도를 높일 수 있다. Attention 알고리즘은 레이블링이 필요하지 않은 unsupervised learning에 의하여 학습될 수 있다.
- [0064] 도 5는 CTC(Connectionist Temporal Classification) 기반 후 처리(post processing)과정을 나타낸다
- [0065] 본 개시의 감지 시스템은 CTC(Connectionist Temporal Classification)를 이용하여 상기 음향 데이터의 각 시간 단계(time-step) 또는 프레임 별(frame wisely)로 각각 레이블링 하여 최종 사건을 감지할 수 있다.
- [0066] 예를 들어, 3 프레임의 음향 데이터가 수신되고 프레임 별 음원 분류 결과가 i) 자동차 타이어 소리, ii) 차량 충격 소음, iii)사람 소리로 결정되었다고 하면, 감지 시스템은 획득된 3프레임의 음향 데이터를 이용하여 최종 사건으로 '교통 사고 발생' 를 출력할 수 있다. 이때, 최종 사건을 판단하는 방법으로 CTC(Connectionist Temporal Classification)가 이용될 수 있다.



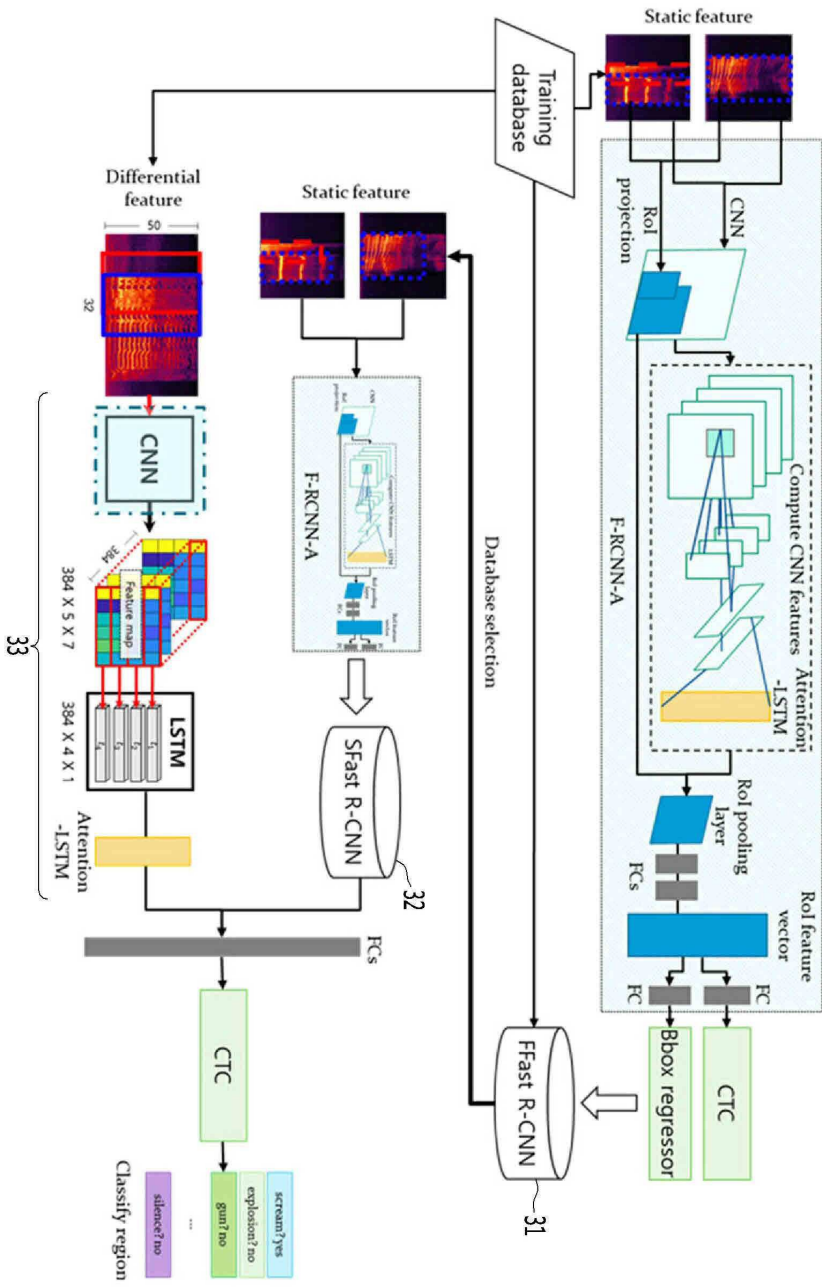
- [0067] 도 6 내지 도 8은 본 개시의 테스트에 따른 성능 결과를 나타낸다.
- [0068] 도 6은 본 개시의 후처리(Post-Processing)에 따른 비교 결과를 나타낸다. 도 6을 참조하면 후처리 존재 여부에 따라서 F1 스코어가 후처리를 한 다음에 더 높은 것을 알 수 있다.
- [0069] 또한, 복합특징값을 사용하고, Attention 알고리즘을 사용하였을 경우 가장 F1스코어가 높은 것을 알 수 있다.
- [0070] 도 7 내지 도 8은 본 개시의 알고리즘의 성능이 정적 특징 및 차등 특징이 포함된 복합특징값을 사용하였을 때 가장 높은 성능을 보인다는 통계수치를 나타낸다.
- [0071] 특히 차등 특징을 사용하는 모델은 정적 기능보다 시끄러운 환경에서 훨씬 뛰어난 성능을 보이는 것을 알 수 있다.
- [0072] 전술한 본 개시는, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있다. 또한, 상기 컴퓨터는 단말기의 프로세서를 포함할 수도 있다.

**도면**

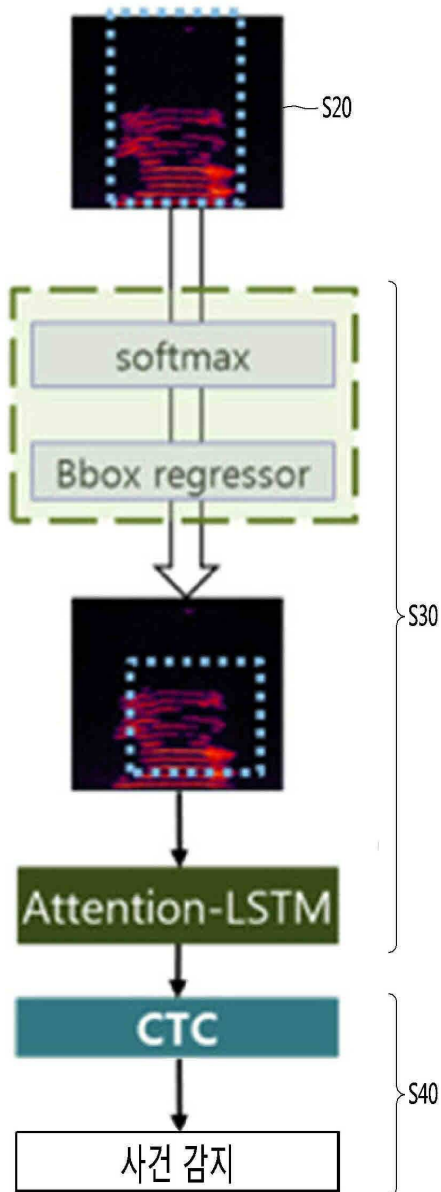
**도면1**



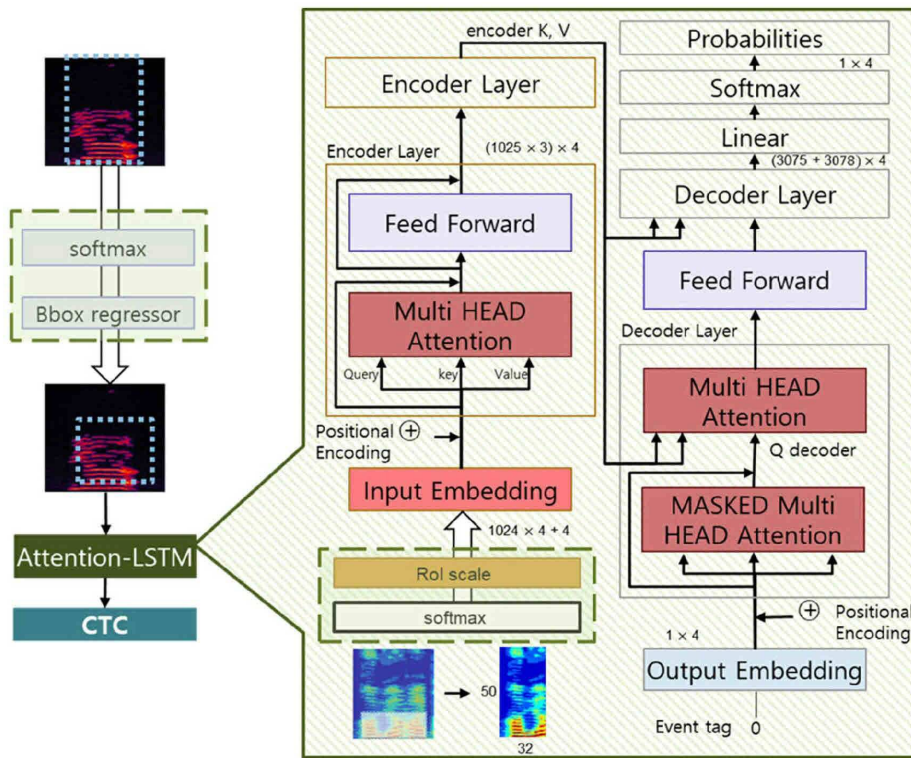
도면2



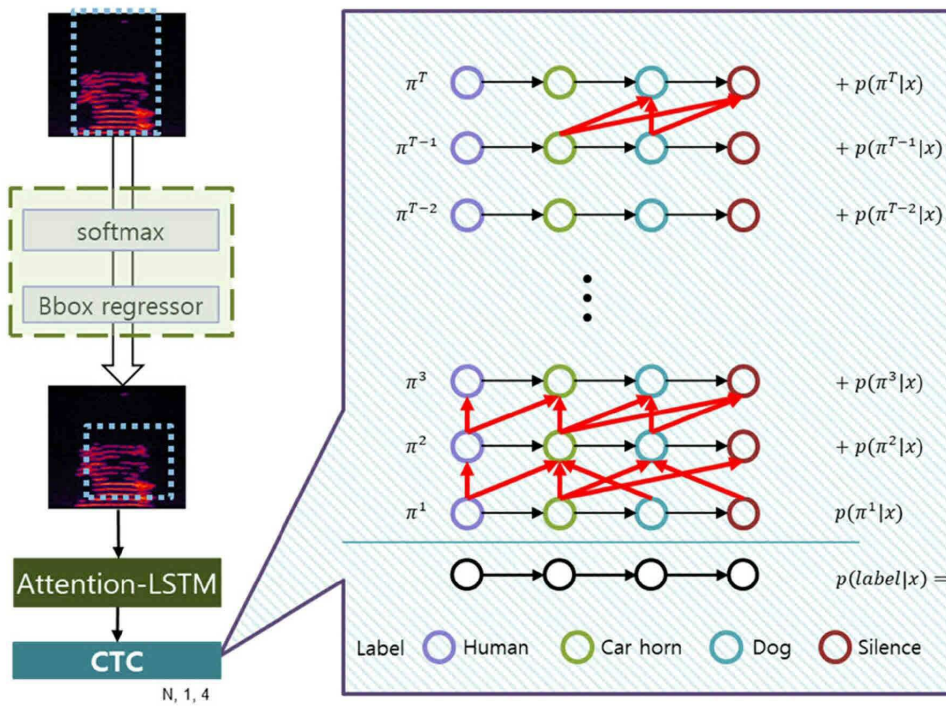
도면3



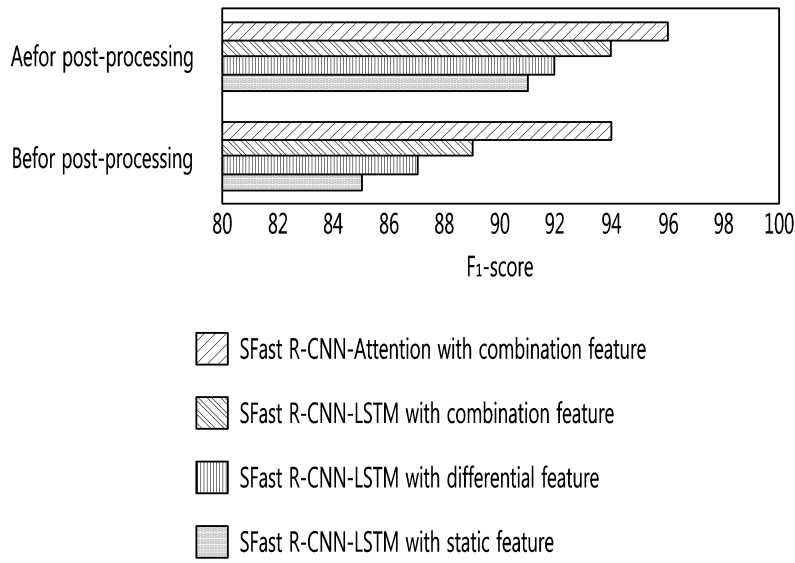
도면4



도면5



도면6



도면7

· Table 1. Performance comparison of F1-score for the fast R-CNN-LSTM AED with static, differential and multi-feature combination according to distances.

Model	Distance					Avg.
	2m	4m	8m	16m	32m	
F-RCNN-LSTM with static feature	0.92	0.89	0.90	0.83	0.73	0.85
F-RCNN-LSTM with differential feature	0.92	0.90	0.90	0.84	0.80	0.87
F-RCNN-LSTM with multi-feature combination	0.93	0.90	0.90	0.88	0.82	0.89

도면8

· Table 2. Performance comparison of F1-score for the fast R-CNN-LSTM AED with i ) static, ii ) differential and iii ) combination feature according to noises that clean (CN), babble(bb), cafe(CF), street(SR), music(MS), and wind(WD).

Model \ Noise	CN	BB	CF	SR	MS	WD	Avg.
F-RCNN-LSTM with static feature	0.87	0.85	0.82	0.83	0.86	0.86	0.85
F-RCNN-LSTM with differential feature	0.82	0.86	0.85	0.86	0.87	0.87	0.86
F-RCNN-LSTM with multi-feature combination	0.89	0.90	0.89	0.89	0.88	0.89	0.89