(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0097490
(43) 공개일자 2013년09월03일

(51) 국제특허분류(Int. Cl.)
G10L 15/20 (2006.01) *G10L 19/00* (2006.01)
(21) 출원번호 10-2012-0019178
(22) 출원일자 2012년02월24일
심사청구일자 없음

(71) 출원인
삼성테크윈 주식회사
경상남도 창원시 성산구 창원대로 1204 (성주동)
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
홍하나
경상남도 창원시 성산구 성주동 28 삼성테크윈
임정은
경상남도 창원시 성산구 성주동 28 삼성테크윈
(뒷면에 계속)
(74) 대리인
리앤목특허법인

전체 청구항 수 : 총 10 항

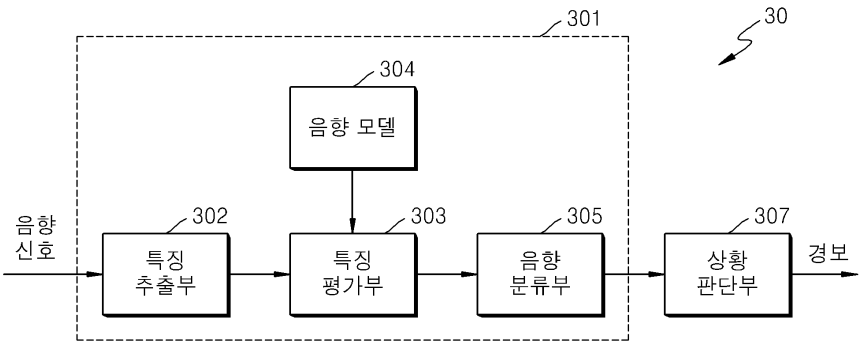
(54) 발명의 명칭 **음향 정보 기반 상황 인식 장치 및 방법**

(57) 요 약

본 발명은 음향 정보 기반 상황 인식 장치 및 방법을 개시한다.

본 발명의 음향 정보 기반 상황 인식 장치는, 입력되는 음향 신호에서 MFCC 특징 및 Timbre 특징을 음향 특징으로 추출하는 특징 추출부와, 복수의 음성 이벤트와 비음성 이벤트를 포함하는 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 기준 음향 특징과 상기 음향 신호의 MFCC 특징 및 Timbre 특징 간의 유사도를 계산하여 상기 음향 특징을 평가하는 특징 평가부와, 상기 음향 신호의 음향 특징과 상기 기준 음향 특징 간의 유사도를 기초로, 상기 음향 신호를 계층적 접근 방식에 의해 상기 복수의 음향 이벤트 중 하나로 분류하는 음향 분류부를 포함할 수 있다.

대 표 도 - 도1



(72) 발명자

고한석

서울특별시 용산구 서빙고동 신동아아파트 16-703

김광윤

서울특별시 성북구 안암동5가 고려대학교자연계캠퍼스 공학관 439호

구본화

서울특별시 성북구 안암동5가 고려대학교자연계캠퍼스 공학관 439호

특허청구의 범위

청구항 1

입력되는 음향 신호에서 MFCC 특징 및 Timbre 특징을 음향 특징으로 추출하는 단계;

복수의 음성 이벤트와 비음성 이벤트를 포함하는 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 기준 음향 특징과 상기 음향 신호의 MFCC 특징 및 Timbre 특징 간의 유사도를 계산하여 상기 음향 특징을 평가하는 단계; 및

상기 음향 신호의 음향 특징과 상기 기준 음향 특징 간의 유사도를 기초로, 상기 음향 신호를 계층적 접근 방식에 의해 상기 복수의 음향 이벤트 중 하나로 분류하는 단계;를 포함하는 음향 정보 기반 상황 인식 방법.

청구항 2

제1항에 있어서, 상기 음향 특징 평가 단계는,

복수의 음향 이벤트 각각에 대해 가우시안 혼합 모델을 통해 미리 획득된 훈련 모델과 상기 음향 신호의 MFCC 특징 간의 우도를 계산하는 단계; 및

상기 음향 신호의 Timbre 특징을 훈련에 의해 미리 획득된 복수의 음향 이벤트 각각의 기준 Timbre 특징과 비교하는 단계;를 포함하는 음향 정보 기반 상황 인식 방법.

청구항 3

제2항에 있어서, 상기 음향 신호 분류 단계는,

최대 우도에 해당하는 음향 이벤트 및 가장 유사한 Timbre 특징을 갖는 음향 이벤트를 선택하여, 상기 음향 신호를 음성 또는 비음성으로 1차 분류하고, 상기 1차 분류된 음향 신호를 복수의 음성 이벤트와 비음성 이벤트 중 하나로 2차 분류하는 단계;를 포함하는 음향 정보 기반 상황 인식 방법.

청구항 4

제1항에 있어서,

일정 시간 동안 입력되는 음향 신호의 분류 결과를 누적하고, 누적된 결과에서 비정상 이벤트의 비율이 임계값 이상이면 비정상 상황으로 판단하여 경보 신호를 출력하는 단계;를 더 포함하는 음향 정보 기반 상황 인식 방법.

청구항 5

제4항에 있어서,

입력되는 영상 신호의 영상 정보를 기초로 비정상 상황의 영상 정보와 비교하고, 비정상 상황으로 판단되면 경보 신호를 출력하는 단계; 및

상기 음향 신호에 대한 경보 신호와 상기 영상 신호에 대한 경보 신호가 모두 출력되면 경보를 발생시키는 단계;를 더 포함하는 음향 정보 기반 상황 인식 방법.

청구항 6

입력되는 음향 신호에서 MFCC 특징 및 Timbre 특징을 음향 특징으로 추출하는 특징 추출부;

복수의 음성 이벤트와 비음성 이벤트를 포함하는 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 기준 음향 특징과 상기 음향 신호의 MFCC 특징 및 Timbre 특징 간의 유사도를 계산하여 상기 음향 특징을 평가하는 특징 평가부; 및

상기 음향 신호의 음향 특징과 상기 기준 음향 특징 간의 유사도를 기초로, 상기 음향 신호를 계층적 접근 방식에 의해 상기 복수의 음향 이벤트 중 하나로 분류하는 음향 분류부;를 포함하는 음향 정보 기반 상황 인식 장치.

청구항 7

제6항에 있어서, 상기 특징 평가부는,

복수의 음향 이벤트 각각에 대해 가우시안 혼합 모델을 통해 미리 획득된 훈련 모델과 상기 음향 신호의 MFCC 특징 간의 우도를 계산하고, 상기 음향 신호의 Timbre 특징을 훈련에 의해 미리 획득된 복수의 음향 이벤트 각각의 기준 Timbre 특징과 비교하는, 음향 정보 기반 상황 인식 장치.

청구항 8

제7항에 있어서, 상기 음향 분류부는,

최대 우도에 해당하는 음향 이벤트를 선택하여, 상기 음향 신호를 음성 또는 비음성으로 1차 분류하는 제1분류기;

최대 우도에 해당하는 음향 이벤트 및 가장 유사한 Timbre 특징을 갖는 음향 이벤트를 선택하여, 음성으로 1차 분류된 음향 신호를 복수의 음성 이벤트 중 하나로 2차 분류하는 제2분류기; 및

최대 우도에 해당하는 음향 이벤트 및 가장 유사한 Timbre 특징을 갖는 음향 이벤트를 선택하여, 비음성으로 1차 분류된 음향 신호를 복수의 비음성 이벤트 중 하나로 2차 분류하는 제3분류기;를 포함하는 음향 정보 기반 상황 인식 장치.

청구항 9

제6항에 있어서,

일정 시간 동안 입력되는 음향 신호의 분류 결과를 누적하고, 누적된 결과에서 비정상 이벤트의 비율이 임계값 이상이면 비정상 상황으로 판단하여 경보 신호를 출력하는 상황 판단부;를 더 포함하는 음향 정보 기반 상황 인식 장치.

청구항 10

제9항에 있어서,

입력되는 영상 신호의 영상 정보를 기초로 비정상 상황의 영상 정보와 비교하고, 비정상 상황으로 판단되면 경보 신호를 출력하는 비디오 모듈; 및

상기 음향 신호에 대한 경보 신호와 상기 영상 신호에 대한 경보 신호가 모두 출력되면 경보를 발생시키는 중앙처리부;를 더 포함하는 음향 정보 기반 상황 인식 장치.

명 세 서

기 술 분 야

[0001] 본 발명은 음향 정보 기반 상황 인식 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 최근 CCTV가 거리 및 빌딩에 보안 및 개인적 안전상의 목적으로 널리 채용되고 있다. 고층 빌딩 내 엘리베이터와 같이 밀폐된 장소에 설치되는 상황 감시 시스템은 통상 감시 카메라의 영상 정보를 이용하여 비정상 이벤트(abnomal event)를 검출하고 기록된 영상 정보는 범죄 증거로 사용된다. 이러한 기존의 상황 감시 시스템은 카메라를 기반으로 이루어져 영상 기반의 상황 분석이 필요하게 된다. 영상 분석만을 이용할 시 특별한 움직임에 대한 정보가 없다면 비정상 상황에 대한 인지가 어렵게 된다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국등록특허 제1070902호는 수집된 음향을 통해 인식된 음성과 사전에 입력된 음성 간의 유사도가 임계값 이상인지 여부에 의해 감시 지역의 이벤트 발생 여부를 판단하는 감시 방법을 개시한다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 음향 정보를 이용하여 비정상 상황의 인식률을 높일 수 있는 음향 인식 방법 및 장치를 제공한다.

과제의 해결 수단

[0005] 본 발명의 바람직한 일 실시예에 따른 음향 정보 기반 상황 인식 방법은, 입력되는 음향 신호에서 MFCC 특징 및 Timbre 특징을 음향 특징으로 추출하는 단계; 복수의 음성 이벤트와 비음성 이벤트를 포함하는 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 기준 음향 특징과 상기 음향 신호의 MFCC 특징 및 Timbre 특징 간의 유사도를 계산하여 상기 음향 특징을 평가하는 단계; 및 상기 음향 신호의 음향 특징과 상기 기준 음향 특징 간의 유사도를 기초로, 상기 음향 신호를 계층적 접근 방식에 의해 상기 복수의 음향 이벤트 중 하나로 분류하는 단계;를 포함할 수 있다.

[0006] 본 발명의 바람직한 일 실시예에 따른 음향 정보 기반 상황 인식 장치는, 입력되는 음향 신호에서 MFCC 특징 및 Timbre 특징을 음향 특징으로 추출하는 특징 추출부; 복수의 음성 이벤트와 비음성 이벤트를 포함하는 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 기준 음향 특징과 상기 음향 신호의 MFCC 특징 및 Timbre 특징 간의 유사도를 계산하여 상기 음향 특징을 평가하는 특징 평가부; 및 상기 음향 신호의 음향 특징과 상기 기준 음향 특징 간의 유사도를 기초로, 상기 음향 신호를 계층적 접근 방식에 의해 상기 복수의 음향 이벤트 중 하나로 분류하는 음향 분류부;를 포함할 수 있다.

발명의 효과

[0007] 본 발명은 MFCC 특징의 단독 사용이 아닌 Timbre 특징을 함께 사용하고, 계층적 분류기의 사용에 의해 좁은 공간 내 음향 이벤트 인식 및 분류 성능을 높일 수 있다. 또한 계층적 분류기의 재구성으로 다양한 환경에 적용이 가능하고, 다양한 이벤트의 음향 정보 데이터를 수집하여 계층적 분류기를 확장할 수 있다.

도면의 간단한 설명

[0008] 도 1은 본 발명의 일 실시예에 따른 음향 인식 장치를 개략적으로 도시한 블록도이다.

도 2는 본 발명의 일 실시예에 따른 MFCC 특징을 획득하는 방법을 설명하는 도면이다.

도 3 및 도 4는 본 발명의 일 실시예에 따른 Timbre 특징을 설명하는 도면이다.

도 5는 본 발명의 일 실시예에 따른 음향 분류부의 구성을 개략적으로 도시한 블록도이다.

도 6은 본 발명의 일 실시예에 따른 음향 정보 기반 상황 인식 방법을 개략적으로 설명하는 흐름도이다.

도 7은 본 발명의 일 실시예에 따른 음향 정보 기반 상황 인식 장치의 구성을 개략적으로 도시한 블록도이다.

도 8은 본 발명의 일 실시예에 따른 영상 신호 기반 비정상 상황을 분류한 예를 도시한 도면이다.

도 9는 본 발명의 실시예의 음향 인식률과 비교예의 음향 인식률을 보여주는 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0009] 이하 본 발명의 바람직한 실시예가 첨부된 도면들을 참조하여 설명될 것이다.

[0010] 도 1은 본 발명의 일 실시예에 따른 음향 인식 장치를 개략적으로 도시한 블록도이다.

[0011] 음향 인식 장치(30)는 음향 신호 분석을 통해 감시 영역에 비정상 상황이 발생되었는지를 판단하고, 비정상 상황으로 판단되면 경보를 발생한다. 도 1을 참조하면, 음향 인식 장치(30)는 음향 이벤트 검출부(301) 및 상황 판단부(307)를 포함할 수 있다. 본 발명의 음향 인식 장치(30)는 엘리베이터, 교도소 등 한정된 공간 안에서 발생 가능한 음향 이벤트들을 계층적으로 분류하여 인식함으로써, 효과적인 상황 인식이 가능하다.

[0012] 음향 이벤트 검출부(301)는 입력되는 음향 신호에서 추출되는 음향 특징을 분석하여, 음향 신호의 이벤트를 계층적으로 인식 및 분류한다. 여기서 이벤트는 비명, 울음, 발걸음, 충돌 등 음향 신호의 종류를 나타낸다. 음

향 이벤트 검출부(301)는 특징 추출부(302), 특징 평가부(303), 음향 모델(304), 음향 분류부(305)를 포함할 수 있다.

[0013] 특징 추출부(302)는 입력되는 음향 신호로부터 음향 특징을 추출한다. 입력되는 음향 신호는 예를 들어, 50ms 단위의 음향 프레임이고, 특징 추출부(302)는 음향 프레임에서 음향 특징을 추출한다. 음향 특징은 MFCC(Mel-frequency cepstral coefficient) 특징 및 Timbre 특징을 포함한다. 특징 추출부(302)는 MFCC 특징 및 Timbre 특징 추출을 차례로 또는 병렬로 수행할 수 있다.

[0014] 도 2는 본 발명의 일 실시예에 따른 MFCC 특징을 획득하는 방법을 설명하는 도면이다. MFCC 특징은 저주파 영역에서의 변화에 상대적으로 민감한 사람의 청각 특성을 반영하여, 고주파 영역에 비해 저주파 영역에서 더 자세히 추출되는 특징 벡터로 음성 인식(speech recognition) 및 다양한 음향 정보를 이용한 인식(acoustic context awarness)에 활용되고 있다.

[0015] 특징 추출부(203)는 시간 영역의 음향 신호를 전처리(pre-processing)하여 고주파 에너지를 부스팅하고(a), 푸리에 변환(FFT)을 취하여 주파수 영역의 스펙트럼을 구한 후(b), 구한 스펙트럼에 대해 멜 스케일(Mel scale)에 맞춘 삼각 필터 뱅크를 대응시켜 각 밴드에서의 크기의 합을 구하고(c), 필터 뱅크 출력값에 로그를 취한 후(d), 이산 코사인 변환(e)을 하여 MFCC 특징 벡터를 획득(f)한다. 이때 MFCC 특징 벡터를 추출하기 위해 윈도우(window)를 적용하게 되는데, 본 발명의 실시예에서는 일예로 50ms 윈도우를 사용하였다. MFCC 특징 추출은 음성인식 기술 분야에서 통상적으로 사용되는 것이므로 보다 구체적인 설술은 생략한다. MFCC 특징은 MFCC 특징 및 MFCC의 시간에 따른 변화량인 Delta 특징을 결합(이하, 'MFCC 특징'으로 통칭함)하여 사용할 수 있다. 본 발명의 일 실시예에서, MFCC 20차 및 Delta 특징 20차를 결합한 특징 벡터를 사용할 수 있다.

[0016] MFCC 특징이 사람의 청각 특성을 반영한 특징이기 때문에 음향 이벤트 인식에서 어느 정도의 성능은 항상 얻을 수 있지만, 엘리베이터 환경에서와 같이 충돌음과 문닫히는 소리처럼 유사한 음향 이벤트들을 구분하여 인식하지 못할 수 있다. 이에 따라 특징 추출부(302)는 성능 향상을 위해 Timber 특징을 음향 특징으로서 추가로 추출한다. Timber 특징은 다른 음원(악기 또는 음성)에 의한 음향 신호가 동일 레벨(level) 및 크기(loudness)를 가지는 경우 이를 구별할 수 있도록 음색 정보를 포함하는 특징이다. Timbre 특징을 MFCC 특징과 결합하여 사용함으로써, 더 강인한 음향 인식 성능을 얻을 수 있다. Timbre 특징은, 밝기(Brightness), 거칠기(Roughness), 영교차율(Zero-crossing rate) 및 롤오프(Roll-off) 특징을 포함할 수 있으며, 본 발명의 실시예는 이에 한정되지 않는다.

[0017] 거칠기(Roughness) 특징은 음향 신호의 정현파 쌍(Sinusoids Pair)의 주파수가 가까움에 따라 발생하는 거친 느낌의 정도를 표현하는 특징으로, 여러 정현파 쌍의 값들을 합산하여 추출한다. 영교차율(Zero-crossing rate) 특징은 음향 신호가 0과 교차되는 정도를 표현하는 특징이다. 롤오프(Roll-off) 특징 및 밝기(Brightness) 특징은 음향 신호에서 고주파의 양을 추정하기 위한 특징으로, 비명, 충돌 소리 등의 비정상 음향과 정상 음향의 주파수 대역에서의 에너지 분포의 차이를 반영한다. 롤오프(Roll-off) 특징은 도 3에 도시된 바와 같이 음향 신호의 전체 주파수 영역에서의 에너지 분포가 저주파에 밀집된 정도를 나타내는 특징으로, 전체 에너지의 소정 비율에 해당하는 주파수(도 3의 예에서는 약 2200Hz)로 표현될 수 있다. 밝기(Brightness) 특징은 도 4에 도시된 바와 같이 음향 신호에서 특정 주파수(도 4의 예에서는 1500Hz) 이상의 고주파 대역에서의 에너지 분포 정도를 나타내는 특징이다.

[0018] 특징 평가부(303)는 음향 신호의 음향 특징과 기준 음향 특징 간의 유사도를 계산하여 음향 특징을 평가한다. 기준 음향 특징은 정의된 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 특징이다. 음향 이벤트는 복수의 음성 이벤트와 복수의 비음성 이벤트를 포함한다.

[0019] 음향 모델(304)은 기준 음향 특징의 데이터베이스로서, 기준 MFCC 특징과 기준 Timbre 특징을 포함한다.

[0020] 기준 Timbre 특징은 훈련 음향 데이터베이스(Database)에서 추출된 Timbre 특징이다. 기준 MFCC 특징은, 훈련 음향 데이터베이스(Database)에서 추출된 MFCC 특징(X)을 이용하여, 하기 식 (1)의 가우시안 혼합 모델(GMM)의 평균(m_i)과 공분산 행렬(R_i) 및 가중치(αi)를 충분히 업데이트 하여 모델링된 음향 이벤트 모델(λ_E)이다. 여기서, k는 혼합되는 가우시안 분포들의 개수를 의미하고, d는 각각의 가우시안 분포의 차수, 즉 MFCC 특징(X)의 차수를 나타낸다.

$$f(x)=\sum_{i=1}^{k}\alpha_i\frac{1}{\sqrt{(2\pi)^d\left|\mathbf{R}_i\right|}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\mathbf{R}_i^{-1}(\mathbf{x}-\mathbf{m}_i)}$$

[0021] ...(1)

[0022] 음향 분류부(305)는 음향 신호의 음향 특징과 기준 음향 특징 간의 유사도를 기초로, 음향 신호를 계층적 접근(hierarchical approch) 방식에 의해 복수의 음향 이벤트 중 하나로 분류한다.

[0023] 일반적으로 발생할 수 있는 음향은 사람의 목에서 발생하는 소리인 음성(vocal)과 그 외의 소리인 비음성(non-vocal), 두 개의 대분류로 나눌 수 있다. 엘리베이터를 예로 들면, 음성은 대화(converstion), 비명(scream), 울음(crying), 안내(announcement) 등의 이벤트로 분류할 수 있고, 비음성은 문개폐(door opening/closing), 발걸음(footsteps), 빈 엘리베이터(empty elevator), 알람벨(alarm bell), 충돌(crash) 등의 이벤트로 분류할 수 있다. 음성 및 비음성 이벤트는 전술된 종류에 한정되지 않고, 감시 영역 및 시스템 설계에 따라 다양하게 설정될 수 있음은 물론이다.

[0024] 음향 분류부(305)는 음향 신호를 먼저 음성과 비음성으로 분류하고, 음성인 경우 복수의 음성 이벤트 중 하나로 분류하고, 비음성인 경우 비음성 이벤트 중 하나로 분류한다.

[0025] 도 5는 본 발명의 일 실시예에 따른 음향 분류부(305)의 구성을 개략적으로 도시한 블록도이다. 음향 분류부(305)는 제1분류기(315), 제2분류기(335), 및 제3분류기(355)를 포함할 수 있다. 각 분류기는 최적화된 음향 특징을 이용하며, 시스템 설치 환경에 따라 재구성이 가능하고, 분류기의 구성에 따라 여러 장소에 적용가능하고 비음성에서도 다양한 이벤트들로 세분화할 수 있다. 또한 각 분류기는 설치 환경에 따라 개별적으로 이벤트 학습을 통해 세분화하여 구성 가능하다.

[0026] 제1분류기(315)는 음향 신호의 MFCC 특징(X)을 기초로 음향 신호를 음성/비음성으로 분류한다. 제1분류기(315)는 하기 식 (2)와 같이 음향 신호의 MFCC 특징(X)의 우도가 가장 큰 음향 이벤트 모델을 선택하여, 선택된 음향 이벤트 모델에 의해 음성 또는 비음성으로 결정한다. 제1분류기(315)는 음성으로 판단되면 음향 신호를 제2분류기(335)로 출력하고, 비음성으로 판단되면 음향 신호를 제3분류기(355)로 출력한다.

$$\hat{E}=\arg\max_{E}P(X\mid\lambda_E)$$

[0027] ...(2)

[0028] 제2분류기(335)는 음성으로 분류된 음향 신호를 음향 신호의 MFCC 특징(X)과 Timbre 특징을 기초로 음성 이벤트들 중 하나로 분류한다. 이때 이용되는 Timbre 특징은 밝기 특징일 수 있다. 제2분류기(335)는 상기 식 (2)와 같이 최대 우도를 갖는 음향 이벤트 모델(λ_E)에 대응하는 음향 이벤트를 찾고, 음향 신호의 밝기 특징과 가장 유사한 기준 밝기 특징을 갖는 음향 이벤트를 찾음으로써, 음향 신호를 음성 이벤트들 중 하나로 분류한다. 이때 제2분류기(335)는 비음성 이벤트의 기준 음향 특징들은 제외하고, 음성 이벤트에 해당하는 음향 이벤트 모델(λ_E)들 간에 계산된 우도와 음성 이벤트들의 기준 밝기 특징과의 비교 결과만을 이용할 수 있다.

[0029] 예를 들어, 최대 우도를 갖는 음향 이벤트 모델(λ_E)이 비명 모델 및 울음 모델인 경우, 음향 신호는 비명과 울음으로 중복 분류될 수 있다. 그러나, 가장 유사한 기준 밝기 특징을 갖는 음성 이벤트가 비명인 경우, 최종적으로 음향 신호를 비명으로 분류할 수 있다.

[0030] 제3분류기(355)는 비음성으로 분류된 음향 신호를 음향 신호의 MFCC 특징(X)과 Timbre 특징을 기초로 비음성 이벤트들 중 하나로 분류한다. 이때 이용되는 Timbre 특징은 거칠기/영교차율/롤오프 특징일 수 있다. 제3분류기(355)는 상기 식 (2)와 같이 최대 우도를 갖는 음향 이벤트 모델(λ_E)을 찾고, 음향 신호의 적어도 하나의 거칠기/영교차율/롤오프 특징과 가장 유사한 적어도 하나의 기준 거칠기/영교차율/롤오프 특징을 갖는 음향 이벤트를 찾음으로써, 음향 신호를 비음성 이벤트들 중 하나로 분류한다. 이때 제3분류기(355)는 음성 이벤트의 기준 음향 특징들은 제외하고, 비음성 이벤트에 해당하는 음향 이벤트 모델(λ_E)들 간에 계산된 우도와 비음성 이벤트들의 기준 거칠기/영교차율/롤오프 특징과의 비교 결과만을 이용할 수 있다.

[0031] 예를 들어, 최대 우도를 갖는 음향 이벤트 모델(λ_E)이 충돌 모델 및 문개폐 모델인 경우, 음향 신호는 충돌과

문개폐로 중복 분류될 수 있다. 그러나, 가장 유사한 적어도 하나의 기준 거칠기/영교차율/롤오프 특징을 갖는 비음성 이벤트가 충돌인 경우, 최종적으로 음향 신호를 충돌로 분류할 수 있다.

[0032] 제2분류기(335) 및 제3분류기(355)는 분류를 위해 MFCC 특징 비교에 의한 우도를 먼저 이용하고, 다음으로 Timbre 특징 비교 결과를 이용할 수 있으며, 그 반대 순서도 가능하며, 본 발명은 특별히 순서에 제한을 두지 않는다.

[0033] 상기 실시예에서는 특징 평가부(303)와 음향 분류부(305)를 별도로 구비하고 있으나, 다른 실시예로서, 음향 분류부(305)의 제1분류기(315), 제2분류기(335), 및 제3분류기(355)가 특징 평가부(303)의 기능을 함께 수행하도록 구현할 수 있다. 이 경우, 제1분류기(315), 제2분류기(335), 및 제3분류기(355) 각각이 훈련에 의한 가우시안 혼합 모델을 구비하여 MFCC 특징의 우도를 계산하고, 최대 우도를 갖는 음향 이벤트 모델을 찾을 수 있다. 또한 제2분류기(335)는 Timbre 특징 중 밝기 특징을 추가로 고려하고, 제3분류기(355)는 Timbre 특징 중 거칠기/영교차율/롤오프 특징을 추가로 고려하여 음향 신호의 이벤트를 분류할 수 있다.

[0034] 상황 판단부(307)는 일정시간 동안 음향 이벤트 분류 결과(인식 결과)를 누적하여, 누적된 이벤트 분류 결과 중에서 비정상 상황(abnomal event)에서 발생할 수 있는 음향 이벤트(이하, '비정상 이벤트'라 함)의 비율이 식(3)과 같이 임계값(threshold) 이상이면, 현재 감시 영역, 예를 들어 엘리베이터 내의 상황을 비정상이라 판단하고 경보 신호를 출력한다. 전술된 음향 이벤트를 예로 들면, 비명, 울음, 충돌이 비정상 이벤트로 분류되고, 나머지는 정상 이벤트로 분류된다.

[0035] $$\frac{\text{비정상 이벤트검출횟수}}{\text{전체이벤트검출횟수}} \geq \text{임계값} \quad \ldots(3)$$

[0036] 도 6은 본 발명의 일 실시예에 따른 음향 정보 기반의 비정상 상황 인식 방법을 개략적으로 설명하는 흐름도이다.

[0037] 음향 인식 장치는 음향 신호가 입력되면, 음향 신호에서 음향 특징을 추출한다(S61). 입력되는 음향 신호는 소정 시간 단위로 샘플링된 음향 프레임이다. 음향 특징은 MFCC 특징 및 Timbre 특징을 포함한다. Timbre 특징으로는 밝기/거칠기/영교차율/롤오프 특징을 포함할 수 있다.

[0038] 음향 인식 장치는 기준 음향 특징과 음향 신호의 음향 특징 간의 유사도를 계산하여 음향 특징을 평가한다(S63). 기준 음향 특징은 복수의 음향 이벤트에 대하여 훈련에 의해 미리 획득된 특징으로, 음향 인식 장치는 기준 음향 특징을 데이터베이스로 구축하여 구비한다. 음향 이벤트는 복수의 음성 이벤트와 복수의 비음성 이벤트를 포함한다. 보다 구체적으로, 음향 인식 장치는 가우시안 혼합 모델을 통해 복수의 음향 이벤트 각각에 대해 미리 획득된 훈련 모델(음향 이벤트 모델)과 음향 신호의 MFCC 특징 간의 우도를 계산한다. 그리고, 음향 인식 장치는 훈련에 의해 미리 획득된 복수의 음향 이벤트 각각의 기준 Timbre 특징을 음향 신호의 Timbre 특징과 비교한다. 음향 인식 장치는 복수의 음향 이벤트를 정의하고, 기준 음향 특징을 미리 데이터베이스로 구축하고 있다.

[0039] 음향 인식 장치는 음향 신호의 음향 특징과 기준 음향 특징 간의 유사도를 기초로, 음향 신호를 계층적 접근 방식에 의해 복수의 음향 이벤트 중 하나로 분류한다. 보다 구체적으로, 음향 인식 장치는 최대 우도에 해당하는 음향 이벤트를 선택하여, 상기 음향 신호를 음성 또는 비음성으로 1차 분류한다(S65).

[0040] 음향 인식 장치는 1차 분류된 음향 신호에 대해, 최대 우도에 해당하는 음향 이벤트 및 가장 유사한 Timbre 특징을 갖는 음향 이벤트를 선택하여, 복수의 음성 이벤트 및 비음성 이벤트들 중 하나로 2차 분류한다(S67). 음향 인식 장치는 음성으로 1차 분류된 음향 신호를, 최대 우도에 해당하는 음성 이벤트 및 가장 유사한 Timbre 특징을 갖는 음성 이벤트를 선택하여 복수의 음성 이벤트들 중 하나의 음성 이벤트로 분류한다. 이때 Timbre 특징으로 밝기 특징이 이용될 수 있다. 음향 인식 장치는 비음성으로 1차 분류된 음향 신호를, 최대 우도에 해당하는 비음성 이벤트 및 가장 유사한 Timbre 특징을 갖는 비음성 이벤트를 선택하여 복수의 비음성 이벤트들 중 하나의 비음성 이벤트로 분류한다. 이때 Timbre 특징으로 거칠기/영교차율/롤오프 특징이 이용될 수 있다.

[0041] 음향 인식 장치는 일정 시간 동안 입력되는 음향 신호의 분류 결과를 누적하고, 누적된 결과에서 비정상 이벤트의 비율이 임계값 이상이면 비정상 상황으로 판단하여 경보 신호를 출력한다(S69).

[0042] 도 7은 본 발명의 일 실시예에 따른 음향 정보 기반 상황 인식 장치의 구성을 개략적으로 도시한 블록도이다.

[0043] 도 7을 참조하면, 음향 정보 기반 상황 인식 장치(1)는 음향 센서(2), 오디오 모듈(3), 카메라(4), 비디오 모듈(5), 중앙처리부(6), 및 출력 장치(7)를 포함한다.

[0044] 음향 센서(2)는 감시 영역에서 발생되는 음향 신호를 수집한다. 음향 센서(2)는 공동주택이나 빌딩에서 엘리베이터 내부, 계단, 지하 주차장, 노인정, 놀이터, 산책로 등과 같은 밀폐 되고 인적이 드문 장소 등에 설치된다. 음향 센서(2)는 마이크 구조에 따라 다이나믹 마이크, 콘덴서 마이크, 리본 마이크 등이 사용될 수 있고, 지향성에 따라 지향성 마이크, 무지향성 마이크, 초지향성 마이크 등이 사용될 수 있다.

[0045] 오디오 모듈(3)은 음향 신호로부터 음향 특징을 추출하여 비정상 상황을 판단한다. 오디오 모듈(3)은 소프트웨어 및/또는 하드웨어에 의해 독립된 모듈로 구현되어, 도 1의 음향 인식 장치(30)의 기능을 수행할 수 있다. 오디오 모듈(3)은 음향 신호 분석을 통해 감시 영역에 비정상 상황이 발생되었는지를 판단하고, 비정상 상황으로 판단되면 경보 신호를 발생한다. 오디오 모듈(3)의 음향 인식 및 비정상 상황 인식 방법은 도 1 내지 도 6을 참조로 설명하였으므로, 상세한 설명은 생략하겠다.

[0046] 카메라(4)는 디지털 및 아날로그 방식의 카메라로, 감시 영역의 영상을 촬영한다. 음향 센서(2)와 카메라(4)는 각각 개별적으로, 또는 음향 센서(2)가 카메라(4)에 내장되는 등의 일체로 설치될 수 있으며, 감시 영역에 하나 이상의 개수로 분산 배치될 수 있다.

[0047] 비디오 모듈(5)은 오디오 모듈(3)과 동일한 공간의 모니터링을 통해 오디오 모듈(3)과 독립적으로 비정상 상황을 감지한다. 비디오 모듈(5)은 영상 신호를 처리 및 분석하여 미리 설정된 비정상 상황에 해당하는지를 판단하고, 비정상 상황으로 판단되면 경보 신호를 발생한다. 비정상 상황은 카메라 템퍼링(camera tempering), 및 영상 내 지정 영역에 물체 검출이 있는 경우를 포함할 수 있다. 도 8은 비디오 모듈(5)에서 판단하는 카메라 템퍼링에 의한 비정상 상황을 분류한 예를 도시하고 있다. 도 8을 참조하면, 카메라 화면 가려짐(a), 카메라 각도 틀어짐(b), 카메라 초점 틀어짐(c), 카메라 화면에 스프레이 분사(d) 등을 비정상 상황으로 분류할 수 있다.

[0048] 중앙처리부(6)는 오디오 모듈(3)에서 판단된 비정상 상황 판단 결과와 비디오 모듈(5)에서 판단된 비정상 상황 판단 결과를 수신하여, 최종적으로 비정상 상황 여부를 판단하고, 출력 장치(7)를 통해 경고 메시지 송출 및/또는 경보음을 발생시킨다. 출력 장치(7)는 디스플레이 및 스피커를 포함할 수 있다. 일 예로, 중앙처리부(6)는 오디오 모듈(3) 및 비디오 모듈(5) 중 하나로부터 경보 신호를 수신하면, 운영자에게 상황 확인이 필요함을 알려준다. 그리고, 중앙처리부(6)는 오디오 모듈(3) 및 비디오 모듈(5) 모두에서 경보 신호를 수신하면 경보를 발생한다. 중앙처리부(6)는 오디오 모듈(3) 및 비디오 모듈(5) 중 적어도 하나에서 경보 신호를 수신하면, 해당 음향 신호 및 영상 신호를 저장한다.

[0049] 중앙처리부(6)는 비정상 상황으로 판단되면 경보를 발생하고, 유선 또는 무선 통신망을 포함하는 네트워크를 통해 통합관리서버 혹은 보안 관제시스템을 연동시켜 즉각적인 대응을 제공할 수 있다.

[0050] 도 9는 본 발명의 실시예의 음향 인식률과 비교예의 음향 인식률을 보여주는 그래프이다. 도 9를 참조하면, MFCC 특징만을 이용한 음향 인식률에 비해, MFCC 특징과 Timbre 특징을 함께 이용한 음향 인식률이 더 우수함을 알 수 있다. 또한 MFCC 특징의 차수가 클수록 음향 인식률이 더 우수함을 알 수 있다.

[0051] 본 발명의 실시예는 한정된 공간 내의 환경에서 발생할 수 있는 음향 이벤트를 효과적인 특징 추출을 통해 모델링하고, 훈련 모델을 이용하여 음향 신호를 계층적으로 분류하여 인식함으로써 강인한 인식 성능을 가질 수 있다. 또한 본 발명의 실시예는 프레임 단위의 인식 결과를 일정 시간 동안 누적하여 누적된 결과 중 비정상 상황에서 발생할 수 있는 음향의 비율이 큰 경우에는 현재 상황을 비정상 상황이라 판단함으로써 오판단을 줄일 수 있다.

[0052] 본 발명의 음향 인식 장치 및 음향 인식 방법은 감시시스템이나 음향 정보를 이용한 지능형 서비스에 활용될 수 있다. 예를 들어, 입력 음향 신호를 분석하고 미리 훈련된 데이터와의 비교를 통해 인식하여 특정 소리의 발생 원인을 찾아내거나, 환경음 인식을 통해 현재 위치를 파악하거나, 총소리나 비명소리 등을 인식하는 감시시스템에 활용될 수 있다.

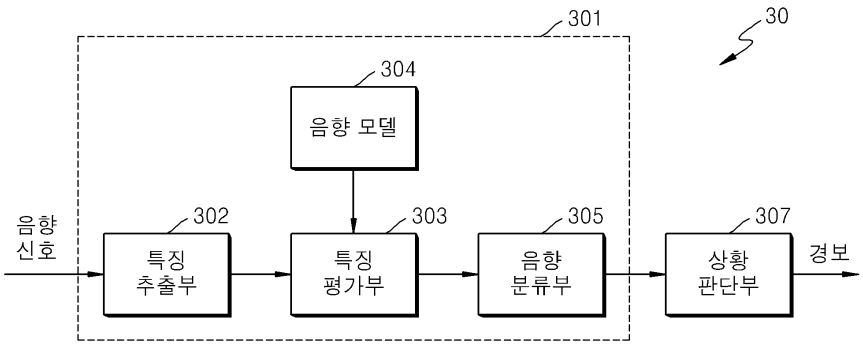
[0053] 야외 환경에서는 영상 정보를 이용한 감시의 경우에 조명 변화 등에 따라 강인한 시스템을 만드는데 어려움이 있으나, 본 발명의 음향 인식 장치 및 음향 인식 방법이 적용된 음향 기반 상황 인식 시스템은, 야외 환경의 음향을 계층적으로 분류하여 비정상 상황을 인식함으로써 강인한 감시 시스템을 구성할 수 있다. 또한 인식하고자 하는 음향 이벤트의 종류나 환경이 바뀌게 되더라도, 일부 음향 특징 추출과 가우시안 혼합 모델의 파라메터

값들만 변화를 주면, 전체 시스템 구성은 동일하게 적용이 가능하기 때문에 다양하게 활용이 가능하다.

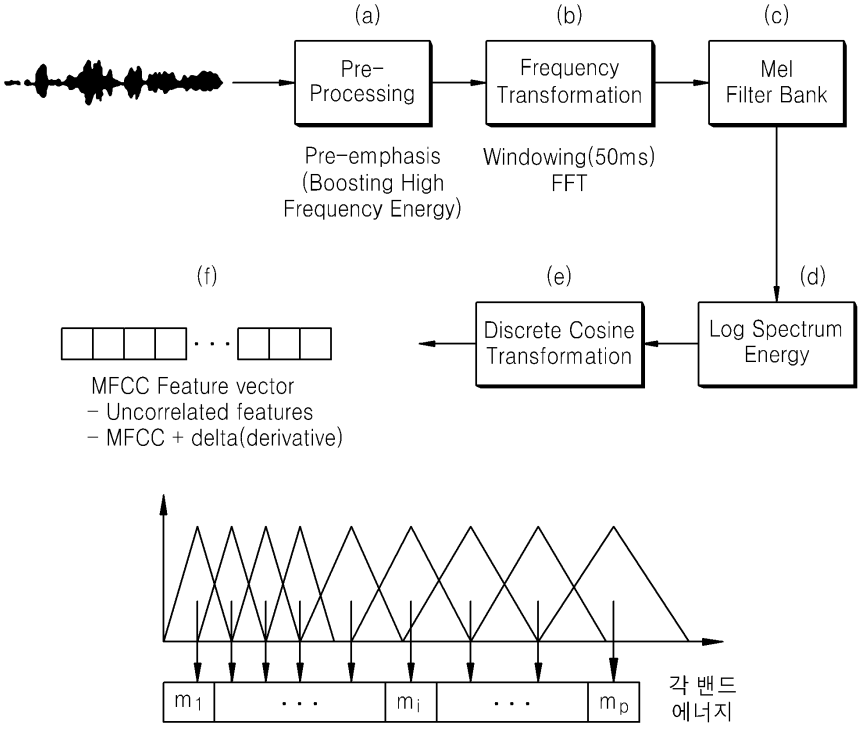
[0054] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

도면

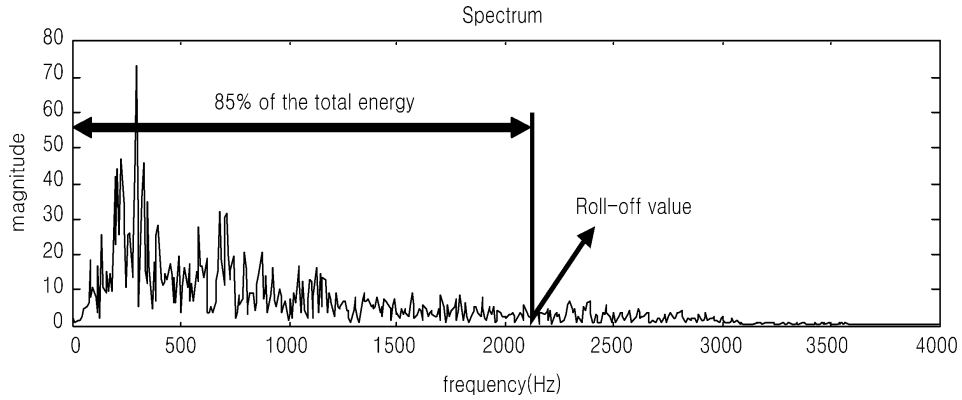
도면1



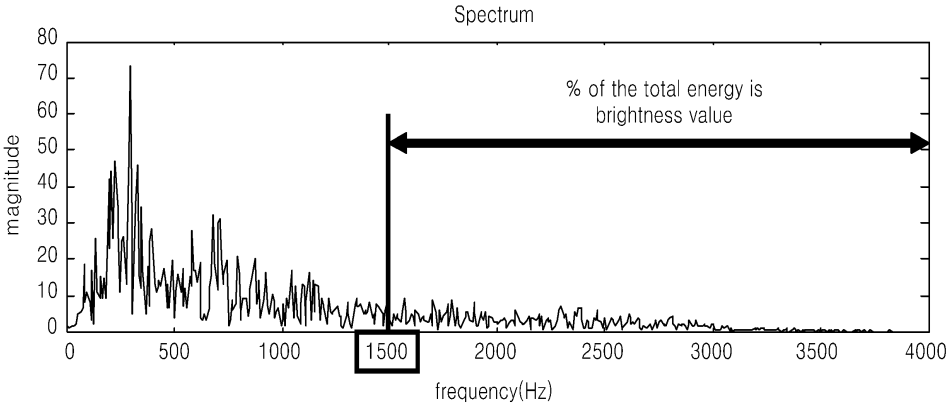
도면2



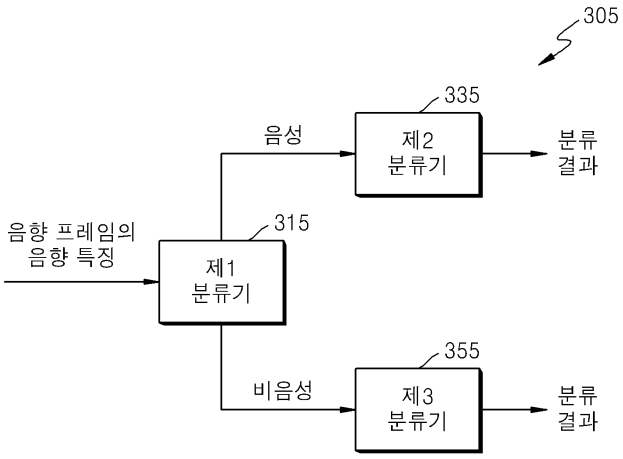
도면3


Spectrum
85% of the total energy
Roll-off value
magnitude
frequency(Hz)
0
10
20
30
40
50
60
70
80
0
500
1000
1500
2000
2500
3000
3500
4000

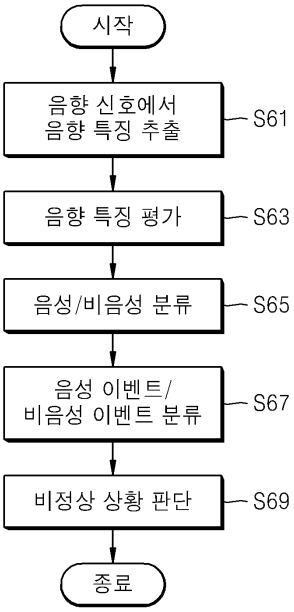
도면4


Spectrum
% of the total energy is
brightness value
magnitude
frequency(Hz)
0
10
20
30
40
50
60
70
80
0
500
1000
1500
2000
2500
3000
3500
4000

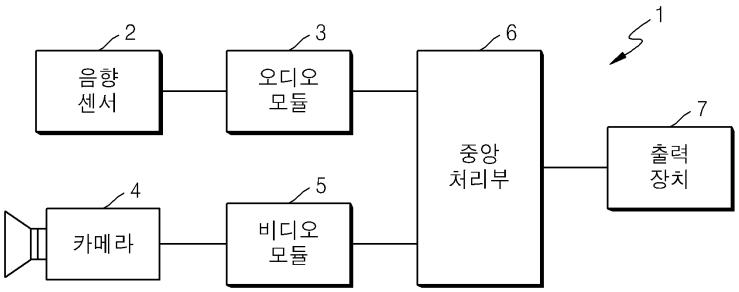
도면5


305
335
음성
제2
분류기
분류
결과
음향 프레임의
음향 특징
315
제1
분류기
355
비음성
제3
분류기
분류
결과

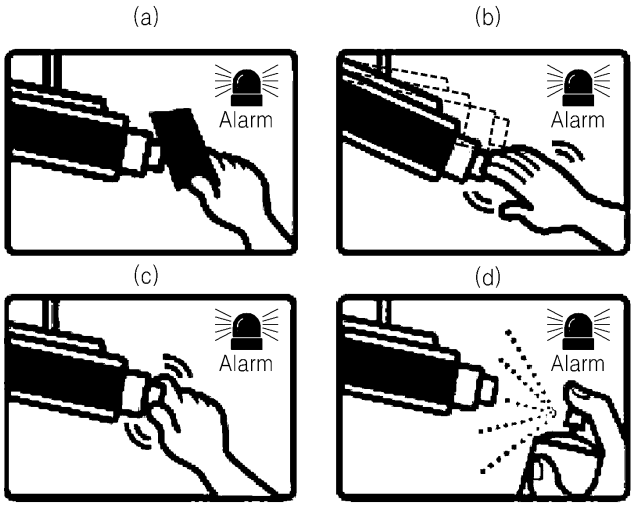
도면6


시작
음향 신호에서 음향 특징 추출
S61
음향 특징 평가
S63
음성/비음성 분류
S65
음성 이벤트/ 비음성 이벤트 분류
S67
비정상 상황 판단
S69
종료

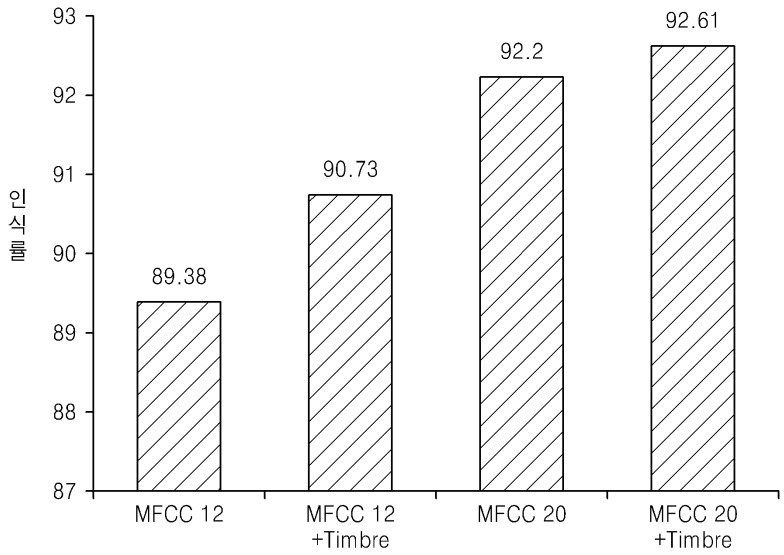
도면7


1
2
음향 센서
3
오디오 모듈
6
중앙 처리부
7
출력 장치
4
카메라
5
비디오 모듈

도면8


(a)
Alarm
(b)
Alarm
(c)
Alarm
(d)
Alarm

도면9


93
92
91
90
89
88
87
인식률
89.38
90.73
92.2
92.61
MFCC 12
MFCC 12 +Timbre
MFCC 20
MFCC 20 +Timbre