



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0100743
(43) 공개일자 2014년08월18일

(51) 국제특허분류(Int. Cl.)
G10L 25/57 (2013.01) G10L 15/25 (2013.01)
G10L 15/28 (2006.01)
(21) 출원번호 10-2013-0013854
(22) 출원일자 2013년02월07일
심사청구일자 2013년02월07일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울 서초구 잠원로 150, 5동 903호 (잠원동, 잠원한신아파트)
제창수
서울 서대문구 통일로25길 30, 103동 701호 (홍제동, 한양아파트)
(74) 대리인
특허법인충현

전체 청구항 수 : 총 34 항

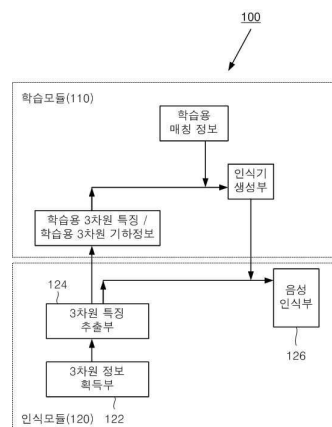
(54) 발명의 명칭 3차원 기하정보를 이용한 음성 인식 방법 및 이를 이용한 음성 인식 시스템

(57) 요약

본 발명은 3차원 기하정보를 이용하여 음성 인식하는 방법 및 시스템에 관한 것이다. 상기 음성 인식 시스템은 학습 모듈 및 인식 모듈을 구비하여, 학습 모듈은 학습용 3차원 기하정보 또는 이들로부터 추출된 학습용 3차원 특징을 이용하여 인식기를 생성한다. 상기 인식 모듈은 음성에 연관되거나 종속된 물리적 대상로부터 획득한 3차원 기하정보 또는 이로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성 인식하는 것을 특징으로 한다.

본 발명에 따른 음성 인식 시스템 및 방법은 발화시의 입술 및 주변, 또는 인체의 임의의 하나 이상의 영역에 대한 3차원 기하정보를 이용하여 음성 인식하는 것을 특징으로 하며, 더 나아가 발화시의 2차원 특징 또는 음향 특징과 3차원 기하정보 또는 3차원 특징을 결합하여 음성 인식하거나, 발화시의 2차원 특징 또는 음향 특징에 의한 인식 결과와 3차원 기하정보 또는 3차원 특징에 의한 인식 결과를 결합하여 최종적으로 음성 인식함으로써 음성 인식의 정확도를 향상시키는 것을 특징으로 한다.

대표도 - 도1



(72) 발명자

김비호

전남 여수시 도원로 140-1, (안산동)

김민욱

경기 고양시 일산동구 호수로 606, B동 904호 (장
항동, 코오롱레이크폴리스I)

특허청구의 범위

청구항 1

학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈;

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

상기 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 2

학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈;

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

상기 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부;

상기 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 3

학습용 2차원 영상으로부터 추출된 학습용 2차원 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈;

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

상기 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부;

상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부; 및

상기 추출된 2차원 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 4

학습용 음향 신호로부터 추출된 학습용 음향 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을

결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈;

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

상기 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부;

외부로부터 음향 신호를 입력받는 음향 신호 입력부;

상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부; 및

상기 3차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 5

학습용 음향 신호들로부터 추출된 학습용 음향 특징, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징 및 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈;

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

상기 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부;

상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부

외부로부터 음향 신호를 입력받는 음향 신호 입력부;

상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부; 및

상기 3차원 특징, 2차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 3차원 정보 획득부가 획득하는 상기 물리적 대상에 대한 3차원 기하정보는, 상기 물리적 대상에 대한 3차원 좌표, 거리, 높이, 디스패리티(disparity) 중 적어도 하나 이상을 포함하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 7

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 3차원 정보 획득부는,

외부로부터 상기 물리적 대상에 대한 3차원 기하정보를 입력받는 3차원 정보 입력부와 상기 물리적 대상에 대한 3차원 정보를 추정할 수 있는 3차원 정보 추정부중 하나로 구성되며,

상기 3차원 정보 추정부는 거리 센서 및 깊이 센서 중 하나 이상으로 구성되는 것을 특징으로 하는 3차원 기하

정보를 이용한 음성 인식 시스템.

청구항 8

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 음성에 연관되거나 종속된 물리적 대상은 인체의 일부 또는 인체의 움직임을 모방한 기계의 일부이거나, 인체 또는 인체의 움직임을 모방한 기계에 착용된 의복의 일부인 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 9

제3항 내지 제5항 중 어느 한 항에 있어서, 서로 다른 2개 이상의 특징에 대하여 딥 러닝(deep learning)하여 서로 다른 2개 이상의 특징을 결합시키는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 10

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 학습용 매칭 정보는 인식기의 입력 데이터와 출력 데이터의 대응 관계에 대한 정보를 포함하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 11

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 이용하여 음성을 인식하는 제1 인식 장치;

상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치;

제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제2 인식 장치는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 12

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치;

외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치;

제1 인식 장치의 인식 결과와 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제3 인식 장치는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 13

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치;

상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치;

외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치;

제1 인식 장치의 인식 결과, 제2 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제2 인식 장치는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈; 을 구비하고,

상기 제3 인식 장치는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 14

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제1 인식 장치;

외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치;

제1 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특

징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제3 인식 장치는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 15

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치;

상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제2 인식 장치;

제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제2 인식 장치는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 2차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 2차원 특징과 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 16

음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제1 인식 장치;

상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치;

제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

상기 제1 인식 장치는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 추출된 3차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

상기 제2 인식 장치는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 상기 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈;

을 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 17

제11항 내지 제16항 중 어느 한 항에 있어서, 상기 제1 인식 모듈은,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부;

상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부;

상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 음성 인식부;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 18

제17항에 있어서, 상기 3차원 정보 획득부가 획득하는 상기 물리적 대상에 대한 3차원 기하정보는, 상기 물리적 대상에 대한 3차원 좌표, 거리, 높이, 디스패리티(disparity) 중 적어도 하나 이상을 포함하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 19

제17항에 있어서, 상기 3차원 정보 획득부는,

외부로부터 상기 물리적 대상에 대한 3차원 기하정보를 입력받는 3차원 정보 입력부와 상기 물리적 대상에 대한 3차원 정보를 추정할 수 있는 3차원 정보 추정부중 하나로 구성되며,

상기 3차원 정보 추정부는 거리 센서 및 깊이 센서 중 하나 이상으로 구성되는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 20

제11항 내지 제16항 중 어느 한 항에 있어서, 상기 물리적 대상은 인체의 일부 또는 인체의 움직임을 모방한 기계의 일부이거나, 인체 또는 인체의 움직임을 모방한 기계에 착용된 의복의 일부인 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 21

제11항 내지 제16항 중 어느 한 항에 있어서, 인식 결합 장치는 둘 이상의 인식 결과들을 사전 설정된 가중치를 부과한 후 결합하여 통합 인식 결과를 생성하여 출력하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 22

제11항 내지 제16항 중 어느 한 항에 있어서, 상기 학습용 매칭 정보는 인식기의 입력 데이터와 출력 데이터의 대응 관계에 대한 정보를 포함하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 23

(a) 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계;

(b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,

상기 인식 단계는,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계;
 상기 획득된 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 단계;
 를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 24

(a) 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계;
 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,
 상기 인식 단계는,
 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계;
 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계;
 상기 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 단계;
 를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 25

(a) 학습용 2차원 영상으로부터 추출된 학습용 2차원 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계;
 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,
 상기 인식 단계는,
 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계;
 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계;
 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계; 및
 상기 추출된 2차원 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;
 를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 26

(a) 학습용 음향 신호로부터 추출된 학습용 음향 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계;
 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,
 상기 인식 단계는,
 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계;
 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계;
 외부로부터 음향 신호를 입력받는 단계;

상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 단계; 및

상기 3차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 27

(a) 학습용 음향 신호들로부터 추출된 학습용 음향 특징, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징 및 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계;

(b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,

상기 인식 단계는,

상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계;

상기 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 단계;

상기 종속된 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계;

외부로부터 음향 신호를 입력받는 단계;

상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 단계; 및

상기 3차원 특징, 2차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 28

(a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 이용하여 음성을 인식하는 제1 인식 단계;

(b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계;

(c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

상기 제2 인식 단계는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 29

(a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계;

- (b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계;
- (c) 제1 인식 단계의 인식 결과와 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계;를 구비하고,

상기 제3 인식 단계는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 30

- (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계;

- (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계;

- (c) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계;

- (d) 제1 인식 단계의 인식 결과, 제2 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계;를 구비하고,

상기 제2 인식 단계는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

상기 제3 인식 단계는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 시스템.

청구항 31

- (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대

상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제1 인식 단계;

(b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계;

(c) 제1 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

상기 제3 인식 단계는,

학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계;

외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 32

(a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계;

(b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제2 인식 단계;

(c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

상기 제2 인식 단계는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 2차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 2차원 특징과 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 33

(a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제1 인식 단계;

(b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단

계;

(c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

상기 제1 인식 단계는,

학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계;

상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 추출된 3차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

상기 제2 인식 단계는,

학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계;

상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 상기 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계;

를 구비하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

청구항 34

제28항 내지 제33항 중 어느 한 항에 있어서, 상기 인식 결합 단계는 둘 이상의 인식 결과들을 사전 설정된 가중치를 부과한 후 결합하여 통합 인식 결과를 생성하여 출력하는 것을 특징으로 하는 3차원 기하정보를 이용한 음성 인식 방법.

명세서

기술분야

[0001] 본 발명은 3차원 기하정보를 이용한 음성 인식 시스템 및 방법에 관한 것으로서, 더욱 구체적으로는 발화자의 입술 및 주변, 또는 인체의 특정 부분에 대한 3차원 기하정보를 이용하여 음성 인식함으로써 보다 정확하게 음성 인식을 수행하는 음성 인식 시스템 및 방법에 관한 것이다.

배경기술

[0002] 음성 인식(speech recognition)은 주로 음향 기반으로 이루어져 왔다. 그러나, 잡음이 큰 환경이나 청각 기능이 부족한 경우에는 입술, 혀 등의 외관 또는 그 영상으로부터 음성을 추정하는 방법이 사용되어 왔다. 또한, 음성 인식의 정확도를 높이기 위하여, 멀티 모달(multi-modal) 기반 연구, 특히 시청각 혼합 연구가 진행되어 왔다.[Matthews, Iain, et al. "Extraction of visual features for lipreading." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.2 (2002): 198-213].

[0003] 주변 음향잡음이 많은 실외, 공장 또는 자동차 운전 환경 등에서는 음향잡음에 무관한 영상 정보를 이용하는 것이 적합하다.

[0004] 종래의 영상에 기반한 시각적 음성인식 방법은 발성자의 입술 모양에 대한 2차원 영상으로부터 추출한 2차원 특징 정보만을 이용하여 음성 인식을 하고 있다. 하지만, 발화시 입술 및 그 주변의 기하학적 변화는 2차원 기하학적 변화에 국한되지 아니하며, 일반적으로 발화시 입술 및 그 주변은 3차원의 기하학적 변화를 겪게 된다.

[0005] 이와 같이, 종래의 기술들은 입술이나 얼굴, 그 외의 다른 인체의 3차원 기하학적 변화를 고려하지 않고 음성 인식함으로써, 음성 인식의 정확도가 떨어지는 문제점이 있었다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 한국공개특허공보 제 10-2010-0062413호
(특허문헌 0002) 한국공개특허공보 제 10-2002-0057046호
(특허문헌 0003) 미국등록특허 US 8,121,347

발명의 내용

해결하려는 과제

- [0007] 전술한 문제점을 해결하기 위한 본 발명의 목적은 발화시의 입술 및 주변, 또는 발화자의 인체의 특정 부분에 대한 3차원 기하정보를 이용하여 음성 인식함으로써 음성 인식의 정확도를 향상시킬 수 있는 음성 인식 시스템 및 음성 인식 방법을 제공하는 것이다.
- [0008] 본 발명의 다른 목적은 발화시의 입술 및 주변, 또는 발화자의 인체의 특정 부분에 대한 3차원 기하정보를 이용하여 음성 인식하는 방법과 기존의 음향 신호를 이용한 음성 인식 방법 및 2차원 영상을 이용하여 음성 인식하는 방법들을 결합하여 음성 인식함으로써 음성 인식의 정확도를 향상시킬 수 있는 음성 인식 시스템 및 음성 인식 방법을 제공하는 것이다.

과제의 해결 수단

- [0009] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 음성 인식 시스템은, 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈; 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,
- [0010] 상기 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;를 구비한다.
- [0011] 본 발명의 제2 특징에 따른 음성 인식 시스템은, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈; 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,
- [0012] 상기 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부; 상기 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;를 구비한다.
- [0013] 본 발명의 제3 특징에 따른 음성 인식 시스템은, 학습용 2차원 영상으로부터 추출된 학습용 2차원 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈; 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,
- [0014] 상기 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부; 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부; 및 상기 추출된 2차원 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;를 구비한다.
- [0015] 본 발명의 제4 특징에 따른 음성 인식 시스템은, 학습용 음향 신호로부터 추출된 학습용 음향 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈; 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡

터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,

- [0016] 상기 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부; 외부로부터 음향 신호를 입력받는 음향 신호 입력부; 상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부; 및 상기 3차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;를 구비한다.
- [0017] 본 발명의 제2 특징에 따른 음성 인식 시스템은, 학습용 음향 신호들로부터 추출된 학습용 음향 특징, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징 및 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 모듈; 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 모듈;을 구비하고,
- [0018] 상기 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부; 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부; 외부로부터 음향 신호를 입력받는 음향 신호 입력부; 상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부; 및 상기 3차원 특징, 2차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부;를 구비한다.
- [0019] 전술한 제1 특징 내지 제5 특징에 따른 음성 인식 시스템에 있어서, 상기 3차원 정보 획득부가 획득하는 상기 물리적 대상에 대한 3차원 기하정보는, 상기 물리적 대상에 대한 3차원 좌표, 거리, 높이, 디스패리티(disparity) 중 적어도 하나 이상을 포함하는 것이 바람직하다.
- [0020] 전술한 제1 특징 내지 제5 특징에 따른 음성 인식 시스템에 있어서, 상기 3차원 정보 획득부는, 외부로부터 상기 물리적 대상에 대한 3차원 기하정보를 입력받는 3차원 정보 입력부와 상기 물리적 대상에 대한 3차원 정보를 추정할 수 있는 3차원 정보 추정부중 하나로 구성되며, 상기 3차원 정보 추정부는 거리 센서 및 깊이 센서 중 하나 이상으로 구성되는 것이 바람직하다.
- [0021] 전술한 제1 특징 내지 제5 특징에 따른 음성 인식 시스템에 있어서, 상기 음성에 연관되거나 종속된 물리적 대상은 인체의 일부 또는 인체의 움직임을 모방한 기계의 일부이거나, 인체 또는 인체의 움직임을 모방한 기계에 착용된 의복의 일부인 것이 바람직하다.
- [0022] 본 발명의 제6 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 이용하여 음성을 인식하는 제1 인식 장치; 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치; 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치;를 구비하고,
- [0023] 상기 제1 인식 장치는, 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기하정보를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈;을 구비하고,
- [0024] 상기 제2 인식 장치는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈;을 구비한다.
- [0025] 본 발명의 제7 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치; 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치; 제1 인식 장치의 인식 결과와 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치;를 구비하고,
- [0026] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기

하정보로부터 3차원 특징을 추출하고, 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

[0027] 상기 제3 인식 장치는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈; 을 구비한다.

[0028] 본 발명의 제8 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치; 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치; 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치; 제1 인식 장치의 인식 결과, 제2 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

[0029] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

[0030] 상기 제2 인식 장치는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈; 을 구비하고,

[0031] 상기 제3 인식 장치는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈; 을 구비한다.

[0032] 본 발명의 제9 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제1 인식 장치; 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 장치; 제1 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

[0033] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,

[0034] 상기 제3 인식 장치는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈; 을 구비한다.

[0035] 본 발명의 제10 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 장치; 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제2 인식 장치; 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,

[0036] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1

인식 모듈; 을 구비하고,

- [0037] 상기 제2 인식 장치는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 2차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 2차원 특징과 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈; 을 구비한다.
- [0038] 본 발명의 제11 특징에 따른 음성 인식 시스템은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제1 인식 장치; 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 장치; 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 장치; 를 구비하고,
- [0039] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 추출된 3차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비하고,
- [0040] 상기 제2 인식 장치는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 상기 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈; 을 구비한다.
- [0041] 전술한 제6 특징 내지 제11 특징에 따른 음성 인식 시스템에 있어서, 상기 제1 인식 모듈은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부; 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부; 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 음성 인식부; 을 구비하는 것이 바람직하다.
- [0042] 본 발명의 제12 특징에 따른 음성 인식 방법은, (a) 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 인식 단계; 를 구비하고,
- [0043] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 획득된 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0044] 본 발명의 제13 특징에 따른 음성 인식 방법은, (a) 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 인식 단계; 를 구비하고,
- [0045] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 상기 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0046] 본 발명의 제13 특징에 따른 음성 인식 방법은, (a) 학습용 2차원 영상으로부터 추출된 학습용 2차원 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계; 를 구비하고,
- [0047] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계; 및 상기 추출된 2차원 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하

고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

- [0048] 본 발명의 제14 특징에 따른 음성 인식 방법은, (a) 학습용 음향 신호로부터 추출된 학습용 음향 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,
- [0049] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 외부로부터 음향 신호를 입력받는 단계; 상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 단계; 및 상기 3차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.
- [0050] 본 발명의 제15 특징에 따른 음성 인식 방법은, (a) 학습용 음향 신호들로부터 추출된 학습용 음향 특징, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징 및 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비하고,
- [0051] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 상기 종속된 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계; 외부로부터 음향 신호를 입력받는 단계; 상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 단계; 및 상기 3차원 특징, 2차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.
- [0052] 본 발명의 제16 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,
- [0053] 상기 제1 인식 단계는, 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,
- [0054] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0055] 본 발명의 제17 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (c) 제1 인식 단계의 인식 결과와 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,
- [0056] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계;를 구비하고,
- [0057] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0058] 본 발명의 제18 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원

영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (d) 제1 인식 단계의 인식 결과, 제2 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

[0059] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

[0060] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

[0061] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

[0062] 본 발명의 제19 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (c) 제1 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

[0063] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

[0064] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

[0065] 본 발명의 제20 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

[0066] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

[0067] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 2차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 2차원 특징과 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

[0068] 본 발명의 제21 특징에 따른 음성 인식 방법은, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을

인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비하고,

[0069] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 추출된 3차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비하고,

[0070] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 상기 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

발명의 효과

[0071] 본 발명에 따른 음성 인식 시스템 및 방법은 발화자의 입술 및 주변, 또는 발화자의 인체의 특정 부분에 대한 3차원 기하정보를 이용하여 음성 인식함으로써 음성 인식의 정확도를 향상시킬 수 있다.

[0072] 또한, 본 발명에 따른 음성 인식 시스템 및 방법은 3차원 기하정보로부터 추출된 3차원 특징과 음향 신호로부터 추출된 음향 특징 또는 2차원 영상으로부터 추출된 2차원 특징을 결합하여 음성 인식함으로써, 음성 인식의 정확도를 더욱 향상시킬 수 있게 한다.

도면의 간단한 설명

[0073] 도 1은 본 발명의 제1 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 2는 3차원 기하학 정보를 획득할 수 있는 다양한 기술들을 예시적으로 도시한 그림으로서, (a)는 구조광 비전 방식 (b)는 전형적인 능동 스테레오 방식, (c)는 (a)와 (b) 방식을 결합한 방식을 나타낸다.

도 3은 본 발명의 제1 실시예에 따른 음성 인식 시스템에 있어서, 음성 인식을 위한 입술 및 주변 영역의 깊이 영상을 예시적으로 도시한 것들로서, 아, 에, 이, 음, 오, 우의 각 음성에 대한 깊이 영상들이다.

도 4는 본 발명의 제1 실시예에 따른 음성 인식 시스템에 있어서, 아, 에, 이, 오, 우의 각 음성에 대하여 3차원 기하정보인 깊이로부터 추출된 특징 벡터들을 예시적으로 도시한 것이다.

도 5는 본 발명의 제2 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 6은 본 발명의 제3 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 7은 본 발명의 제4 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 8은 딥 러닝을 설명하기 위하여 도시한 모식도이다.

도 9는 본 발명의 제5 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 10은 본 발명의 제6 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 11은 본 발명의 제7 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 12는 본 발명의 제8 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 13은 본 발명의 제9 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

도 14는 본 발명의 제10 실시예에 따른 음성 인식 시스템을 도시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0074] 본 발명의 바람직한 실시예들에 따른 음성 인식 시스템 및 방법은 발화자의 입술 및 주변, 또는 인체의 임의의 하나 이상의 영역에 대한 3차원 기하정보를 이용하여 음성 인식하는 것을 특징으로 하며, 더 나아가 발화자의 2

차원 특징 또는 음향 특징과 3차원 기하정보 또는 3차원 특징을 결합하여 음성 인식하거나, 발화시의 2차원 특징 또는 음향 특징에 의한 인식 결과와 3차원 기하정보 또는 3차원 특징에 의한 인식 결과를 결합하여 최종적으로 음성 인식함으로써 음성 인식의 정확도를 향상시키는 것을 특징으로 한다.

- [0075] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예들에 따른 음성 인식 시스템 및 그 방법을 구체적으로 설명한다.
- [0076] < 제1 실시예 : 3차원 기하정보를 이용한 음성 인식 시스템 및 방법>
- [0077] 본 발명의 제1 실시예에 따른 음성 인식 시스템은 발화자의 발화시의 입술 및 주변, 또는 인체의 일부분에 대한 3차원 기하정보를 이용하여 음성 인식하는 것을 특징으로 한다.
- [0078] 도 1은 본 발명의 제1 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 1을 참조하여 본 발명의 제1 실시예에 따른 음성 인식 시스템 및 방법을 구체적으로 설명한다.
- [0079] 도 1을 참조하면, 본 발명의 제1 실시예에 따른 음성 인식 시스템(100)은 학습 모듈(110) 및 인식 모듈(120)로 이루어진다.
- [0080] 상기 학습 모듈(110)은 학습용 3차원 기하정보를 그대로 이용하거나 상기 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징 또는 학습용 3차원 기하 정보와 학습용 매칭 정보를 이용하여 인식기를 생성한다. 상기 학습용 3차원 특징 또는 학습용 기하 정보를 이용하여 인식기를 생성하는 경우, PCA(principal component analysis) 또는 LDA(linear discriminant analysis) 등의 방법을 이용하여 각 특징의 차원을 효과적으로 감소시킬 수 있다. 상기 인식기는 널리 알려져 있는 GMM(Gaussian mixture model), NN(nearest neighbor) 알고리즘, k-NN(k-nearest neighbor) 알고리즘 등을 이용하여 생성 가능하며, 그 외에도 다양한 알고리즘을 사용할 수 있을 것이다.
- [0081] 상기 학습용 매칭 정보는, 사람, 기계 또는 소프트웨어가 생성하는 것으로서, 인식의 입력 데이터와 출력 데이터 사이의 직관적 또는 통계적 대응 관계를 포함하는 것을 특징으로 한다.
- [0082] 상기 인식 모듈(120)은 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 획득하고, 상기 3차원 기하정보 또는 상기 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하여 출력한다. 상기 3차원 기하정보는 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보이다.
- [0083] 상기 음성에 연관되거나(correlated) 종속된(dependent) 물리적 대상은, 인체의 일부 또는 인체의 움직임을 모방한 기계(예, 휴머노이드)의 일부이거나, 인체 또는 인체의 움직임을 모방한 기계에 착용된 의복의 일부인 것을 특징으로 한다(예: 입술, 치아, 혀, 뺨, 턱, 눈, 눈썹, 손, 장갑, 마스크, 휴머노이드의 입 또는 손 등).
- [0084] 상기 물리적 대상에 대한 3차원 기하정보는 상기 물리적 대상의 적어도 하나 이상의 영역의 표면의 3차원 기하정보이거나, 상기 물리적 대상의 적어도 하나 이상의 영역의 경계선 또는 거계곡선을 따라 획득된 3차원 기하정보이거나, 또는 상기 물리적 대상의 적어도 하나 이상의 영역에서 특징적인 하나 이상의 위치에서 얻어진 3차원 기하정보인 것이 바람직하다. 상기 3차원 기하 정보는 3차원 좌표로 표현되거나, 2차원 영상 좌표에 추가하여 3차원 특징을 갖게 하는 거리, 깊이 또는 디스패리티(disparity)의 형태가 될 수 있다. 상기 디스패리티(disparity)는 두 시점(viewpoints)의 영상 차이값을 나타낸다.
- [0085] 본 발명에 따른 음성 인식 시스템 및 음성 인식 방법에 있어서, 상기 3차원 기하정보는 기하학적 보정(geometric calibration) 없이 디스패리티의 형태로 사용되거나, 기하학적 보정을 거친 3차원 좌표 형태로 사용되거나, 3차원 기하정보의 샘플링, 차원 축소(dimension reduction), 제로민(zero-mean) 처리, 분산정규화(variance normalization) 중 하나 이상의 처리를 통해 3차원 특징을 추출하여 사용될 수 있다.
- [0086] 본 명세서의 전반에 걸쳐 기재된 음성에 연관되거나 종속된 물리적 대상, 3차원 기하정보, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하 정보, 학습용 매칭 정보는 전술한 의미와 모두 동일한 의미로 사용되므로, 이후 중복되는 설명은 생략한다.
- [0087] 상기 학습 모듈의 학습용 3차원 특징은 학습용 3차원 기하정보로부터 추출된 3차원 특징이며, 학습용 매칭 정보는 각 3차원 특징들 또는 학습용 3차원 기하정보들에 각각 매칭되는 음성 정보를 의미한다. 따라서, 상기 학습

모듈은 학습용 3차원 기하정보들 또는 이들로부터 추출된 학습용 3차원 특징, 그리고 학습용 매칭 정보들을 이용하여 인식기를 생성한다.

- [0088] 상기 인식 모듈(120)은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부(122), 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부(124), 상기 추출된 3차원 특징 또는 3차원 기하정보를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부(126)를 구비한다.
- [0089] 상기 3차원 정보 획득부(122)는, 외부로부터 상기 물리적 대상에 대한 3차원 기하정보를 입력받는 3차원 정보 입력부를 포함하거나, 상기 물리적 대상에 대한 3차원 기하정보를 직접 추정하는 3차원 기하정보 추정부를 포함하여 구성될 수 있다. 상기 3차원 기하정보 추정부를 포함하는 경우, 상기 3차원 기하정보 추정부는 현존하는 다양한 거리 센서(range sensor)들 또는 깊이 센서(depth sensor)들 중 하나 이상을 포함할 수 있으며(예, 키넥트(Kinect)), 이들의 대표적인 측정 방식은 스테레오 비전(stereo vision), 구조광(structured light) 등이 있다. 도 2는 3차원 기하정보를 획득할 수 있는 다양한 기술들을 예시적으로 도시한 그림으로서, (a)는 구조광 비전 방식 (b)는 전형적인 능동 스테레오 방식, (c)는 (a)와 (b) 방식을 결합한 방식을 나타낸다.
- [0090] 도 3은 본 발명의 바람직한 실시예에 따른 음성 인식 시스템에 있어서, 음성 인식을 위한 입술 및 주변 영역의 깊이 영상을 예시적으로 도시한 것들로서, 아, 에, 이, 음, 오, 우의 각 음성에 대한 깊이 영상들이다.
- [0091] 도 4는 본 발명의 바람직한 실시예에 따른 음성 인식 시스템에 있어서, 아, 에, 이, 오, 우의 각 음성에 대하여 3차원 기하정보인 깊이로부터 추출된 특징들을 예시적으로 도시한 것이다. 도 4에 있어서, 1로 표시된 초록선은 '에', 2로 표시된 파랑선은 '이', 3으로 표시된 마젠타선은 '오', 4로 표시된 빨강선은 '아', 5로 표시된 검정선은 '우'의 음성에 대한 3차원 특징들을 각각 나타낸다.
- [0092] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 학습 모듈과 인식 모듈에 각각 적용되는 방법들로서, 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하거나 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계와, 상기 물리적 대상에 대한 3차원 기하정보 또는 이로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 인식 단계를 구비한다.
- [0093] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계, 상기 획득된 3차원 기하정보 또는 상기 3차원 기하정보로부터 추출된 3차원 특징을 상기 인식기에 적용하여 음성을 인식하는 단계를 구비한다.
- [0094] < 제2 실시예 : 3차원 기하정보와 2차원 특징을 이용한 음성 인식 시스템 및 방법 >
- [0095] 본 발명의 제2 실시예에 따른 음성 인식 시스템 및 음성 인식 방법은 음성에 연관되거나 종속된 물리적 대상에 대한 2차원 영상을 이용하여 음성 인식하는 것을 특징으로 한다.
- [0096] 도 5는 본 발명의 제2 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 5를 참조하여 본 발명의 제2 실시예에 따른 음성 인식 시스템 및 방법을 구체적으로 설명한다.
- [0097] 도 5를 참조하면, 본 발명의 제2 실시예에 따른 음성 인식 시스템(200)은 학습 모듈(210) 및 인식 모듈(220)로 이루어진다. 상기 학습 모듈(210)은 학습용 2차원 특징 및 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 인식기를 생성한다.
- [0098] 상기 인식 모듈(220)은 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하여 출력한다. 상기 인식 모듈이 인식기를 생성하는 방식은 제1 실시예에서의 인식기 생성 방법과 동일하므로, 이후 중복되는 설명은 생략한다.
- [0099] 상기 학습 모듈의 학습용 3차원 특징은 학습용 3차원 기하정보들로부터 추출된 학습용 3차원 특징이며, 학습용 2차원 특징은 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징이며, 학습용 매칭 정보는 학습용 3차원 특징들, 학습용 2차원 특징들 및 학습용 3차원 기하정보들에 매칭되는 음성 정보를 의미한다. 상기 학습 모듈은 학습용 2차원 및 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성한다.
- [0100] 상기 인식 모듈(220)은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부(222), 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부(224), 상기

물리적 대상에 대한 2차원 영상을 획득하는 2차원 영상 획득부(232), 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부(234), 및 상기 추출된 2차원 및 3차원 특징들을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부(226)를 구비한다.

- [0101] 상기 3차원 정보 획득부(222), 3차원 특징 추출부(224)는 제1 실시예의 그것들과 동일하므로 중복되는 설명은 생략한다.
- [0102] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 학습 모듈과 인식 모듈에 각각 적용되는 방법들로서, (a) 학습용 2차원 영상으로부터 추출된 학습용 2차원 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; 및 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 2차원 영상으로부터 추출된 2차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비한다.
- [0103] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계; 및 상기 추출된 2차원 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.
- [0104] < 제3 실시예 : 3차원 기하정보와 음향 특징을 이용한 음성 인식 시스템>
- [0105] 본 발명의 제3 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보 및 음향 신호로부터 추출된 음향 특징을 이용하여 음성 인식하는 것을 특징으로 한다.
- [0106] 도 6은 본 발명의 제3 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 6을 참조하여 본 발명의 제3 실시예에 따른 음성 인식 시스템을 구체적으로 설명한다.
- [0107] 도 6을 참조하면, 본 발명의 제3 실시예에 따른 음성 인식 시스템(300)은 학습 모듈(310) 및 인식 모듈(320)로 이루어진다. 상기 학습 모듈(310)은 학습용 음향 특징 및 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 인식기를 생성한다. 상기 인식 모듈(320)은 상기 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 생성된 특징 벡터를 상기 인식기에 적용하여 음성을 인식하여 출력한다.
- [0108] 상기 학습 모듈의 학습용 3차원 특징은 학습용 3차원 기하정보들로부터 추출된 3차원 특징이며, 학습용 음향 특징은 학습용 음향 신호들로부터 추출된 음향 특징이며, 학습용 매칭 정보는 학습용 3차원 특징들, 학습용 음향 특징들 및 학습용 3차원 기하정보들에 매칭되는 음성 정보를 의미한다. 상기 학습 모듈은 학습용 음향 특징 및 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 생성된 학습용 특징 벡터를 이용하여 인식기를 생성한다.
- [0109] 상기 인식 모듈(320)은, 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부(322), 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부(324), 외부로부터 음향 신호를 입력받는 음향 신호 입력부(332), 상기 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부(334), 및 상기 추출된 음향 특징 및 3차원 특징들을 결합하여 하나의 특징 벡터를 생성하고, 상기 생성된 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부(326)를 구비한다.
- [0110] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 학습 모듈과 인식 모듈에 각각 적용되는 방법들로서, (a) 학습용 음향 신호로부터 추출된 학습용 음향 특징 및 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; 및 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비한다.
- [0111] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 외부로부터 음향 신호를 입력받는 단계; 상기 음향 신호 입력부로부터 입력된 음향 신호

로부터 음향 특징을 추출하는 단계; 및 상기 3차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.

[0112] < 제4 실시예 : 3차원 기하정보, 2차원 특징 및 음향 특징을 이용한 음성 인식 시스템>

[0113] 본 발명의 제4 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상에 대한 2차원 영상, 3차원 기하정보 및 음향 신호를 이용하여 음성 인식하는 것을 특징으로 한다.

[0114] 도 7은 본 발명의 제4 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 7을 참조하여 본 발명의 제4 실시예에 따른 음성 인식 시스템을 구체적으로 설명한다.

[0115] 도 7을 참조하면, 본 발명의 제4 실시예에 따른 음성 인식 시스템(400)은 학습 모듈(410) 및 인식 모듈(420)로 이루어진다. 상기 학습 모듈(410)은 학습용 음향 특징, 학습용 2차원 특징 및 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 인식기를 생성한다. 상기 인식 모듈(420)은 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용함으로써, 음성을 인식하여 출력한다.

[0116] 상기 학습 모듈의 학습용 3차원 특징은 학습용 3차원 기하정보들로부터 추출된 3차원 특징이며, 학습용 2차원 특징은 학습용 2차원 영상으로부터 추출된 2차원 특징이며, 학습용 음향 특징은 학습용 음향 신호들로부터 추출된 음향 특징이며, 학습용 매칭 정보는 학습용 2차원 특징들, 학습용 3차원 특징들, 학습용 음향 특징들 및 학습용 3차원 기하정보들에 매칭되는 음성 정보를 의미한다. 상기 학습 모듈은 학습용 음향 특징, 학습용 2차원 특징 및 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 상기 학습용 매칭 정보를 이용하여 인식기를 생성한다.

[0117] 상기 인식 모듈(420)은, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 3차원 정보 획득부(422), 상기 3차원 정보 획득부에 의해 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 3차원 특징 추출부(424), 상기 물리적 대상에 대한 2차원 영상을 획득하는 2차원 영상 획득부(432), 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 2차원 특징 추출부(434), 외부로부터 음향 신호를 입력받는 음향 신호 입력부(442), 상기 입력된 음향 신호로부터 음향 특징을 추출하는 음향 특징 추출부(444), 및 상기 추출된 음향 특징, 2차원 특징 및 3차원 특징들을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 음성 인식부(426)를 구비한다.

[0118] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 학습 모듈과 인식 모듈에 각각 적용되는 방법들로서, (a) 학습용 음향 신호들로부터 추출된 학습용 음향 특징, 학습용 3차원 기하정보로부터 추출된 학습용 3차원 특징 및 학습용 2차원 영상들로부터 추출된 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 인식기를 생성하는 학습 단계; 및 (b) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 음향 신호로부터 추출된 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 인식 단계;를 구비한다.

[0119] 상기 인식 단계는, 상기 물리적 대상에 대한 3차원 기하정보를 획득하는 단계; 상기 획득된 3차원 기하정보로부터 3차원 특징을 추출하는 단계; 상기 물리적 대상에 대한 2차원 영상을 획득하고, 상기 획득된 2차원 영상으로부터 2차원 특징을 추출하는 단계; 외부로부터 음향 신호를 입력받는 단계; 상기 음향 신호 입력부로 입력된 음향 신호로부터 음향 특징을 추출하는 단계; 및 상기 3차원 특징, 2차원 특징 및 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.

[0120] 전술한 제2 실시예 내지 제4 실시예는 음향 특징, 2차원 특징 및 3차원 특징들 중 적어도 둘 이상을 인식 이전에 결합하여 통합된 하나의 특징 벡터를 구한 뒤, 인식을 수행하는 선결합(early intrgration) 방식을 적용한 것이다. 이러한 선결합 방식은 두 특징을 특징 수준(feature level)에서 통합하는 방법으로서, 영상 및 음향 특징을 추출한 후 두 가지 특징 중 잡음 환경에서도 강인한 특징을 나타낼 수 있는 결합을 찾아 영상 및 음향 통합 특징을 만들어 내는 것이 바람직하다.

- [0121] 이러한 방법은 영상 및 음향에서 잡음환경에 강인한 정보들을 두드러지게 하면서 특징 벡터의 차원을 축소할 수 있는 장점이 있다.
- [0122] 이하, 전술한 제2 실시예 내지 제4 실시예에 있어서 멀티 모달 딥러닝(deep learning) 방법에 의한 음성 및 영상의 통합 특징 획득 과정을 구체적으로 설명한다. 도 8은 딥 러닝을 설명하기 위하여 도시한 모식도이다.
- [0123] 도 8을 참조하면, 음성 및 영상의 특징을 특징 수준에서 결합할 때 딥 러닝을 이용함으로써 보다 효율적인 특징을 획득할 수 있다.
- [0124] 딥 러닝(deep learning)은 러닝의 레이어수가 3개 이상인 러닝 구조에서의 통합된 러닝을 의미한다. 먼저, pretraining 단계에서는 기본 러닝 구조의 러닝을 RBM(restricted Boltzmann machine)을 통해 실시하고, unrolling 단계에서는 deep autoencoder 를 생성한 후, fine tuning 단계를 통해 딥 러닝을 완료한다. 딥 러닝을 하면 PCA나 shallow 러닝을 하는 것보다 component간의 correlation을 보다 효율적으로 기술할 수 있다. 도 8에 도시된 바와 같이, 음향 신호(Audio input)와 영상 신호(Video input)에 기반한 멀티 모달의 경우에는 각 신호에 대한 딥 러닝의 결과들이 통합된 feature 레이어를 형성한다. 이때 딥 러닝을 통해 통합된 representation 의 차원적 효율성이 달성된다. 특히, 두 모달간의 correlation을 보다 효율적으로 기술할 수 있다.
- [0125] < 제5 실시예 : 3차원 기하정보를 이용한 음성 인식 결과와 2차원 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 및 음성 인식 방법 >
- [0126] 본 발명의 제5 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용한 음성 인식 결과와, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용한 음성 인식 결과를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.
- [0127] 도 9는 본 발명의 제5 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 9를 참조하여 본 발명의 제5 실시예에 따른 음성 인식 시스템 및 방법을 구체적으로 설명한다.
- [0128] 도 9를 참조하면, 본 발명의 제5 실시예에 따른 음성 인식 시스템(500)은 제1 인식 장치(510), 제2 인식 장치(520) 및 인식 결합 장치(540)를 구비한다.
- [0129] 상기 제1 인식 장치(510)는 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 것이며, 제2 인식 장치(520)는 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(530)는 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.
- [0130] 상기 제1 인식 장치는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 학습용 3차원 특징 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 및 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈을 구비한다.
- [0131] 상기 제2 인식 장치는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징 및 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈, 및 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈을 구비한다.
- [0132] 상기 인식 결합 장치(530)는 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.
- [0133] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제2 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보를 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비한다.

- [0134] 상기 제1 인식 단계는, 학습용 3차원 기하정보와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0135] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0136] < 제6 실시예 : 3차원 기하정보를 이용한 음성 인식 결과와 음향 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 >
- [0137] 본 발명의 제6 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상의 3차원 기하정보로부터 추출된 3차원 특징을 이용한 음성 인식 결과와, 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용한 음성 인식 결과를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.
- [0138] 도 10은 본 발명의 제6 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 10을 참조하여 본 발명의 제6 실시예에 따른 음성 인식 시스템 및 음성 인식 방법을 구체적으로 설명한다.
- [0139] 도 10을 참조하면, 본 발명의 제6 실시예에 따른 음성 인식 시스템(600)은 제1 인식 장치(610), 제3 인식 장치(630) 및 인식 결합 장치(640)를 구비한다.
- [0140] 상기 제1 인식 장치(610)는 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 것이며, 제3 인식 장치(630)는 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(640)는 제1 인식 장치의 인식 결과와 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.
- [0141] 상기 제1 인식 장치는 제5 실시예의 제1 인식 장치와 동일하므로, 중복된 설명은 생략한다.
- [0142] 상기 제3 인식 장치는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 제3 학습 모듈, 및 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 제3 인식 모듈을 구비한다.
- [0143] 상기 인식 결합 장치(630)는 제1 인식 장치의 인식 결과와 제3 인식 장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.
- [0144] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제3 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (c) 제1 인식 단계의 인식 결과와 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비한다.
- [0145] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0146] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0147] < 제7 실시예 : 3차원 기하정보를 이용한 음성 인식 결과, 2차원 특징을 이용한 음성 인식 결과 및 음향 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 및 방법 >
- [0148] 본 발명의 제7 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상의 3차원 기하정보로부터 추출된 3차원 특징을 이용한 음성 인식 결과, 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용한 음성 인식 결과와, 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용한 음성 인식 결과

를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.

- [0149] 도 11은 본 발명의 제7 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 11을 참조하여 본 발명의 제7 실시예에 따른 음성 인식 시스템을 구체적으로 설명한다.
- [0150] 도 11을 참조하면, 본 발명의 제7 실시예에 따른 음성 인식 시스템(700)은 제1 인식 장치(710), 제2 인식 장치(720), 제3 인식 장치(730) 및 인식 결합 장치(740)를 구비한다.
- [0151] 상기 제1 인식 장치(710)는 상기 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 것이며, 제2 인식 장치(720)는 상기 리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 것이며, 제3 인식 장치(730)는 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(740)는 제1 인식 장치의 인식 결과, 제2 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.
- [0152] 상기 제1 인식 장치, 제2 인식 장치 및 제3 인식 장치는 제5 및 제6 실시예의 그것들과 동일하므로, 중복된 설명은 생략한다.
- [0153] 상기 인식 결합 장치(730)는 제1 인식 장치의 인식 결과, 제2 인식장치의 인식 결과 및 제3 인식 장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.
- [0154] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제2 인식 장치, 제3 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (d) 제1 인식 단계의 인식 결과, 제2 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계를 구비한다.
- [0155] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.
- [0156] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 추출된 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0157] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0158] < 제8 실시예 : 3차원 기하정보와 2차원 특징을 이용한 음성 인식 결과와 음향 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 >
- [0159] 본 발명의 제8 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상의 3차원 기하정보로부터 추출된 3차원 특징 및 음성에 연관되거나 종속된 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용한 음성 인식 결과와, 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용한 음성 인식 결과를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.
- [0160] 도 12는 본 발명의 제8 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 12를 참조하여 본 발명의 제8 실시예에 따른 음성 인식 시스템을 구체적으로 설명한다.
- [0161] 도 12를 참조하면, 본 발명의 제8 실시예에 따른 음성 인식 시스템(800)은 제1 인식 장치(810), 제3 인식 장치(830) 및 인식 결합 장치(840)를 구비한다.

- [0162] 상기 제1 인식 장치(810)는 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 것이며, 제3 인식 장치(830)는 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(840)는 제1 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.
- [0163] 상기 제1 인식 장치(810)는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 학습용 2차원 특징 및 학습용 3차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈, 및 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈을 구비한다.
- [0164] 상기 제3 인식 장치는 제6 실시예의 그것들과 동일하므로, 중복된 설명은 생략한다.
- [0165] 상기 인식 결합 장치(830)는 제1 인식 장치의 인식 결과 및 제3 인식 장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.
- [0166] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제3 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제3 인식 단계; (c) 제1 인식 단계의 인식 결과 및 제3 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비한다.
- [0167] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 2차원 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 추출된 2차원 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0168] 상기 제3 인식 단계는, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 음향 특징과 학습용 매칭 정보를 이용하여 제3 인식기를 생성하는 단계; 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징을 상기 제3 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0169] < 제 9 실시예 : 3차원 기하정보와 음향 특징을 이용한 음성 인식 결과와 2차원 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 및 방법 >
- [0170] 본 발명의 제9 실시예에 따른 음성 인식 시스템은 음성에 연관되거나 종속된 물리적 대상의 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용한 음성 인식 결과와 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용한 음성 인식 결과를 이용한 음성 인식 결과를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.
- [0171] 도 13은 본 발명의 제9 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 13을 참조하여 본 발명의 제9 실시예에 따른 음성 인식 시스템 및 음성 인식 방법을 구체적으로 설명한다.
- [0172] 도 13를 참조하면, 본 발명의 제9 실시예에 따른 음성 인식 시스템(900)은 제1 인식 장치(910), 제2 인식 장치(920) 및 인식 결합 장치(940)를 구비한다.
- [0173] 상기 제1 인식 장치(910)는 상기 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 것이며, 제2 인식 장치(920)는 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(940)는 제1 인식 장치의 인식 결과와 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.
- [0174] 상기 제1 인식 장치(910)는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로

부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징과 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 제1 학습 모듈; 및 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 음향 특징 및 3차원 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 제1 인식 모듈; 을 구비한다.

[0175] 상기 제2 인식 장치는 제5 실시예의 그것들과 동일하므로, 중복된 설명은 생략한다.

[0176] 상기 인식 결합 장치(940)는 제1 인식 장치의 인식 결과 및 제2 인식장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.

[0177] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제2 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비한다.

[0178] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 3차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 추출된 3차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

[0179] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 상기 학습용 2차원 특징과 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 상기 추출된 2차원 특징을 상기 제2 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.

[0180] < 제10 실시예 : 3차원 기하정보를 이용한 음성 인식 결과, 및 2차원 특징과 음향 특징을 이용한 음성 인식 결과를 결합한 음성 인식 시스템 >

[0181] 본 발명의 제10 실시예에 따른 음성 인식 시스템은 발화자의 음성에 연관되거나 종속된 물리적 대상의 3차원 기하정보로부터 추출된 3차원 특징을 이용한 음성 인식 결과, 및 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징과 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용한 음성 인식 결과를 결합하여 최종적으로 음성 인식을 수행하는 것을 특징으로 한다.

[0182] 도 14는 본 발명의 제10 실시예에 따른 음성 인식 시스템을 도시한 블록도이다. 이하, 도 14를 참조하여 본 발명의 제10 실시예에 따른 음성 인식 시스템 및 음성 인식 방법을 구체적으로 설명한다.

[0183] 도 14를 참조하면, 본 발명의 제10 실시예에 따른 음성 인식 시스템(1000)은 제1 인식 장치(1010), 제2 인식 장치(1020), 및 인식 결합 장치(1040)를 구비한다. 상기 제1 인식 장치(1010)는 상기 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 것이며, 제2 인식 장치(1020)는 상기 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징, 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 것이며, 인식 결합 장치(1040)는 제1 인식 장치의 인식 결과 및 제2 인식 장치의 인식 결과를 이용하여 최종적으로 음성을 결정한다.

[0184] 상기 제1 인식 장치는 제5 실시예의 그것과 동일하므로, 중복된 설명은 생략한다.

[0185] 상기 제2 인식 장치(1020)는 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 추출된 학습용 2차원 특징 및 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터 및 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 제2 학습 모듈; 및 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 추출된 2차원 특징과 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 제2 인식 모듈; 을 구비한다.

- [0186] 상기 인식 결합 장치(1040)는 제1 인식 장치의 인식 결과 및 제2 인식장치의 인식 결과를 SNR에 기반한 가중치(weighting factors)로 결합하여 통합 인식 결과를 생성하여 제공하는 것을 특징으로 한다.
- [0187] 본 실시예에 따른 음성 인식 방법은 전술한 음성 인식 시스템의 제1 인식 장치, 제2 인식 장치 및 인식 결합 장치에 각각 적용되는 방법들로서, (a) 음성에 연관되거나 종속된 물리적 대상에 대한 3차원 기하정보로부터 추출된 3차원 특징을 이용하여 음성을 인식하는 제1 인식 단계; (b) 음성에 연관되거나 종속된 물리적 대상에 대한 2차원 영상으로부터 추출된 2차원 특징 및 외부로부터 입력된 음향 신호로부터 추출된 음향 특징을 이용하여 음성을 인식하는 제2 인식 단계; (c) 제1 인식 단계의 인식 결과와 제2 인식 단계의 인식 결과를 이용하여 최종적으로 음성을 결정하는 인식 결합 단계; 를 구비한다.
- [0188] 상기 제1 인식 단계는, 학습용 3차원 기하정보로부터 학습용 3차원 특징을 추출하고, 상기 추출된 학습용 3차원 특징과 학습용 매칭 정보를 이용하여 제1 인식기를 생성하는 단계; 상기 물리적 대상에 대한 3차원 기하정보로부터 3차원 특징을 추출하고, 추출된 3차원 특징을 상기 제1 인식기에 적용하여 음성을 인식하는 단계; 를 구비한다.
- [0189] 상기 제2 인식 단계는, 학습용 2차원 영상으로부터 학습용 2차원 특징을 추출하고, 학습용 음향 신호로부터 학습용 음향 특징을 추출하고, 상기 학습용 2차원 특징 및 상기 학습용 음향 특징을 결합하여 학습용 특징 벡터를 생성하고, 상기 학습용 특징 벡터와 학습용 매칭 정보를 이용하여 제2 인식기를 생성하는 단계; 상기 물리적 대상에 대한 2차원 영상으로부터 2차원 특징을 추출하고, 외부로부터 입력된 음향 신호로부터 음향 특징을 추출하고, 상기 2차원 특징과 상기 음향 특징을 결합하여 하나의 특징 벡터를 생성하고, 상기 특징 벡터를 상기 제2 인식기에 적용하여 음성을 인식하는 단계;를 구비한다.
- [0190] 전술한 제5 내지 제10 실시예는 음향 특징과 영상 특징에 기반한 인식을 각각 수행한 후 두 인식 결과를 SNR에 기반한 적절한 가중치로 결합하여 통합 인식 결과를 얻는 방식으로 후결합(late integration) 방법이다. 이 방법은 시각 및 음성 신호 각각에 잘맞는 인식 방법을 선택하여 인식을 수행할 수 있는 장점이 있다.
- [0191] 이상에서와 같이 본 발명은 단일의 바람직한 실시예를 들어 도시하고 설명하였으나, 본 발명은 상기한 실시예에 한정되지 아니하며 본 발명의 정신을 벗어나지 않는 범위내에서 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 다양한 변형과 수정이 가능할 수 있을 것이다.

산업상 이용가능성

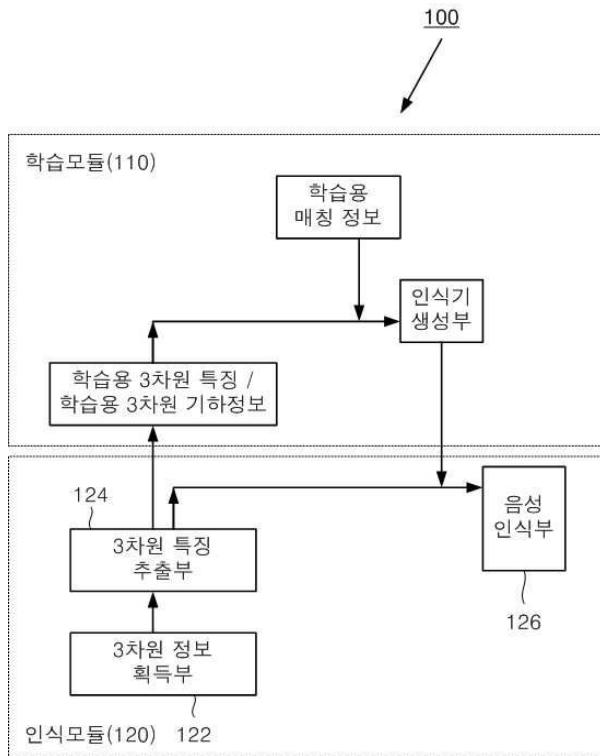
- [0192] 본 발명에 따른 음성 인식 시스템 및 음성 인식 방법은 음성 인식 분야에 널리 사용될 수 있다.

부호의 설명

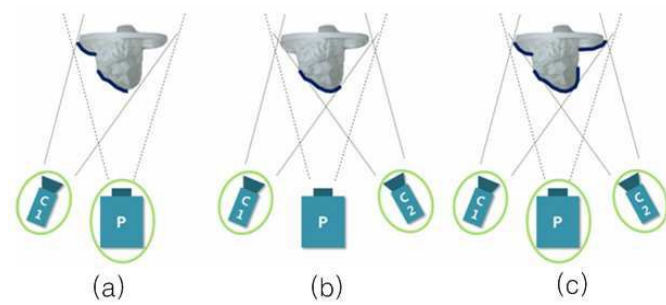
- [0193] 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000:음성 인식 시스템
- 110, 210, 310, 410 : 학습 모듈
- 120, 220, 320, 420 : 인식 모듈
- 122, 222, 322, 422 : 3차원 정보 획득부
- 124, 224, 324, 424 : 3차원 특징 추출부
- 126, 226, 326, 426 : 음성 인식부
- 232 : 2차원 영상 획득부
- 234 : 2차원 특징 추출부
- 510, 610, 710, 810, 910, 1010 : 제1 인식 장치
- 520, 720, 920, 1020 : 제2 인식 장치
- 630, 730, 830 : 제3 인식 장치
- 540, 640, 740, 840, 940, 1040 : 인식 결합 장치

도면

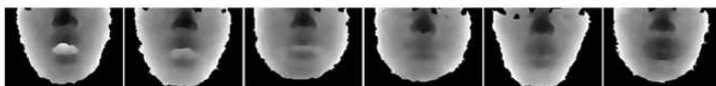
도면1



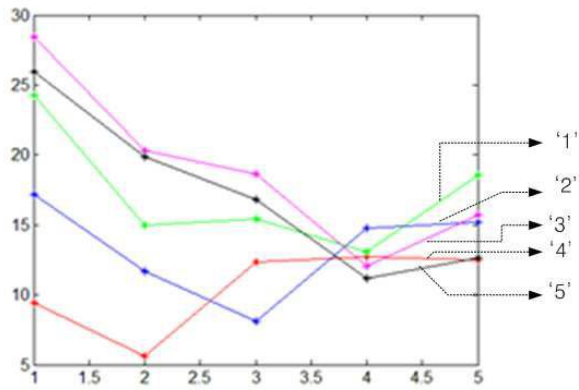
도면2



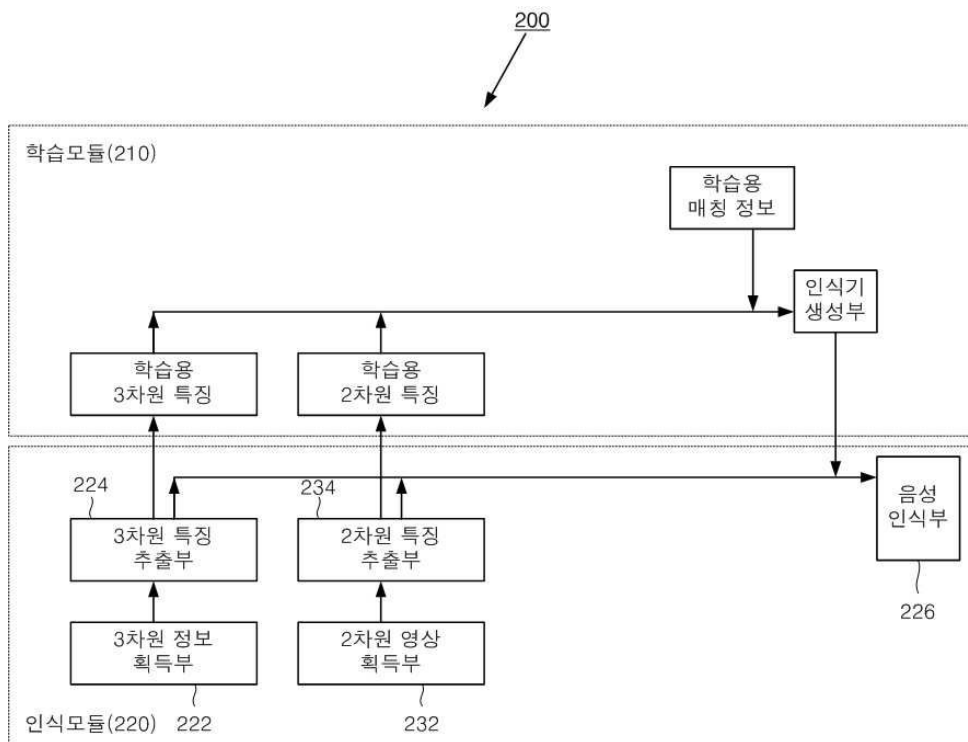
도면3



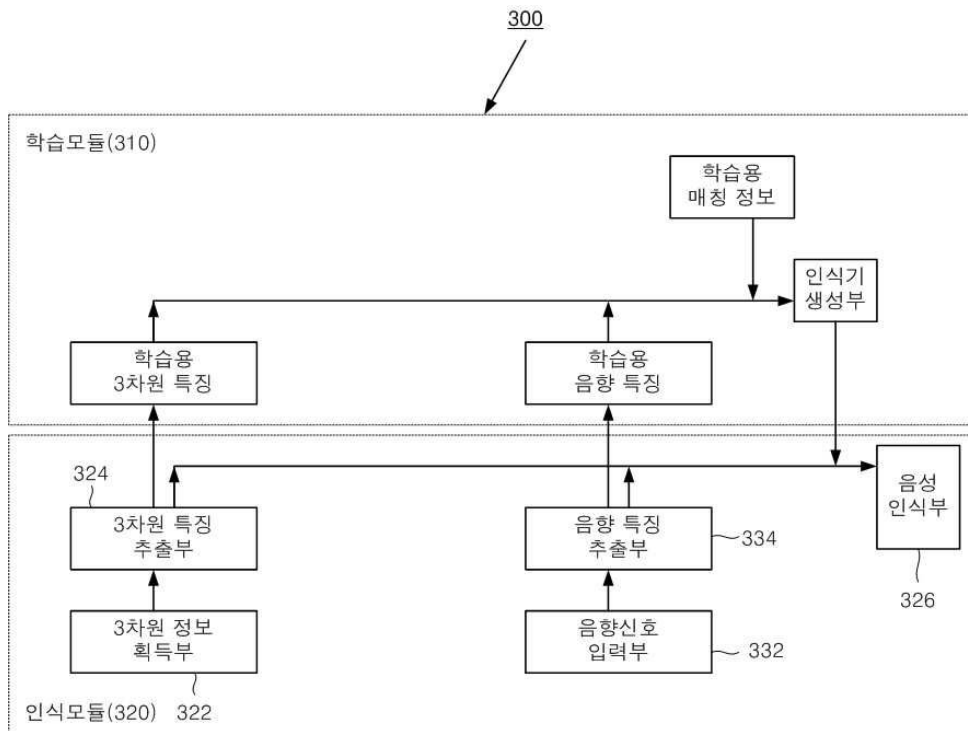
도면4



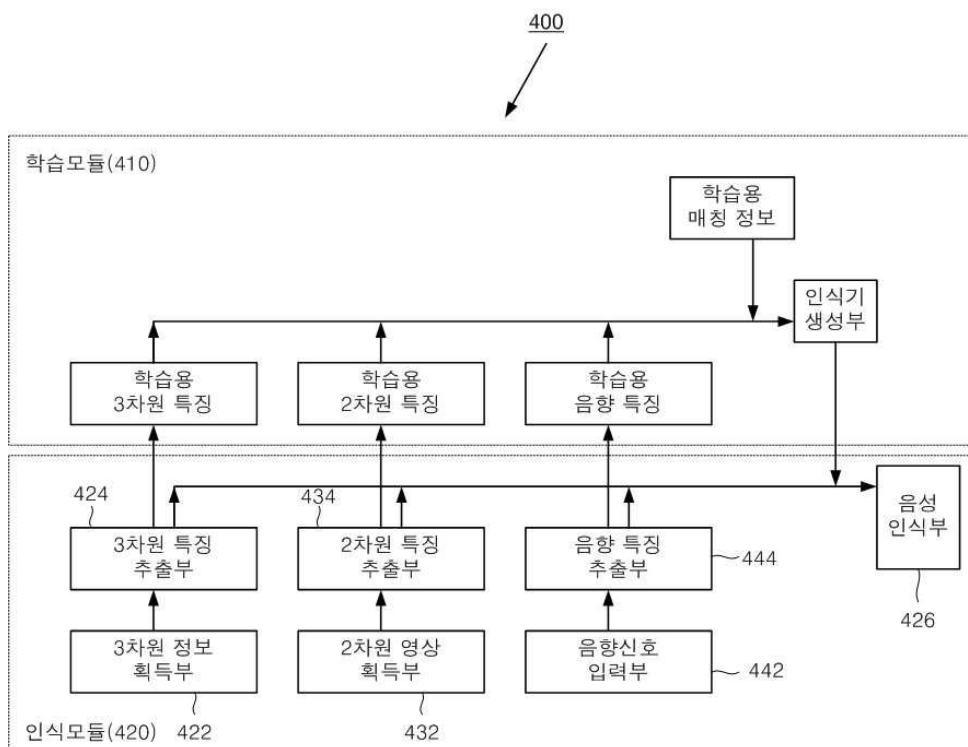
도면5



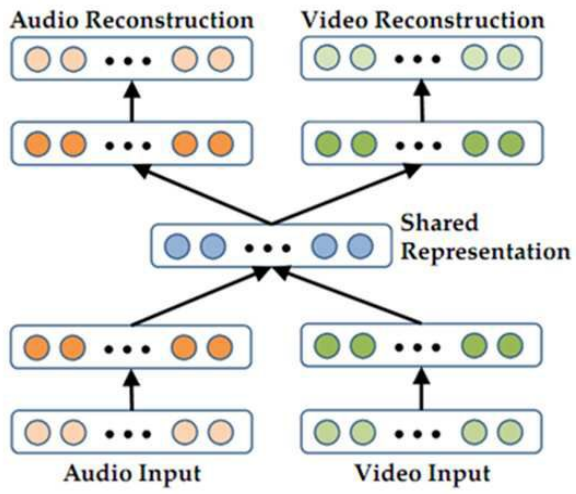
도면6



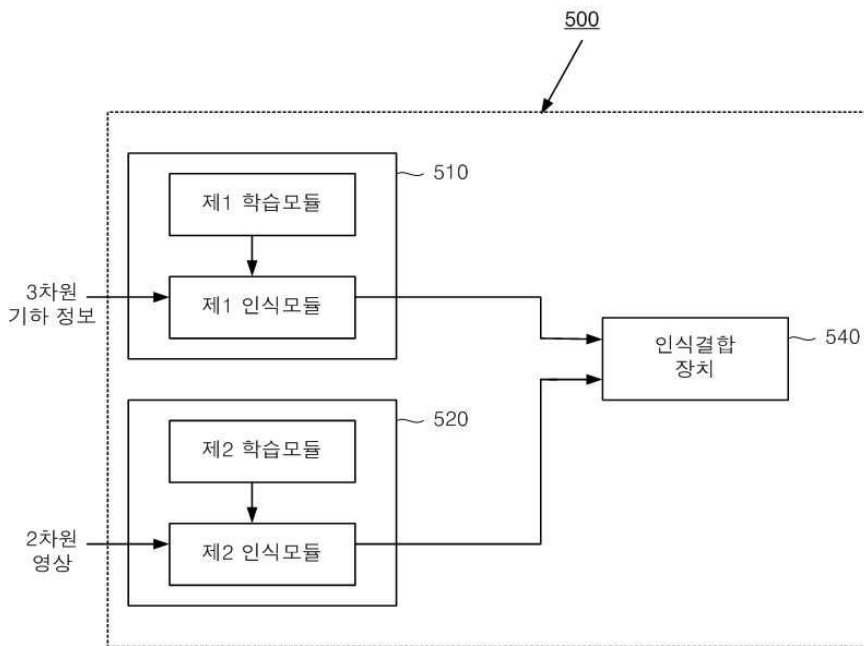
도면7



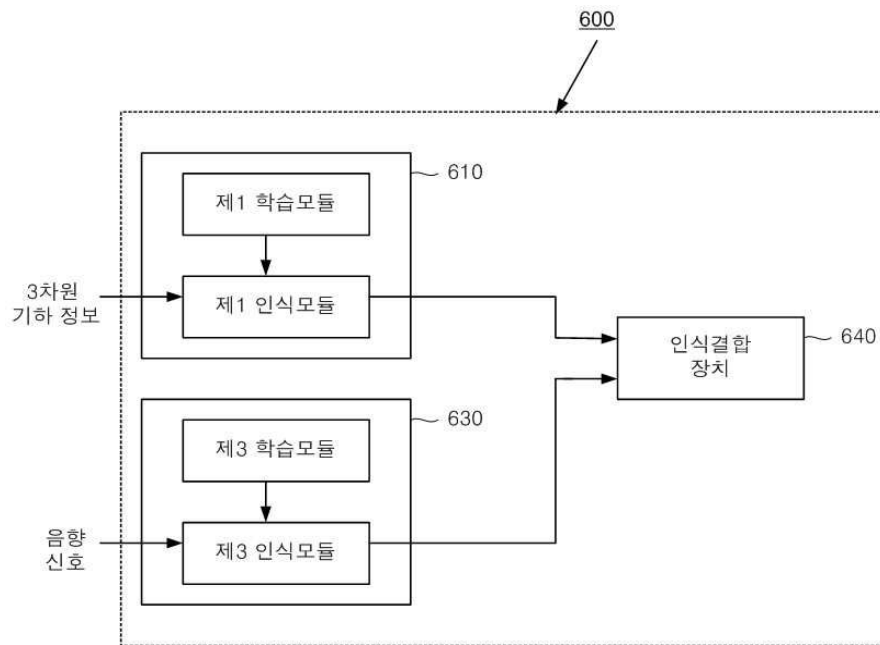
도면8



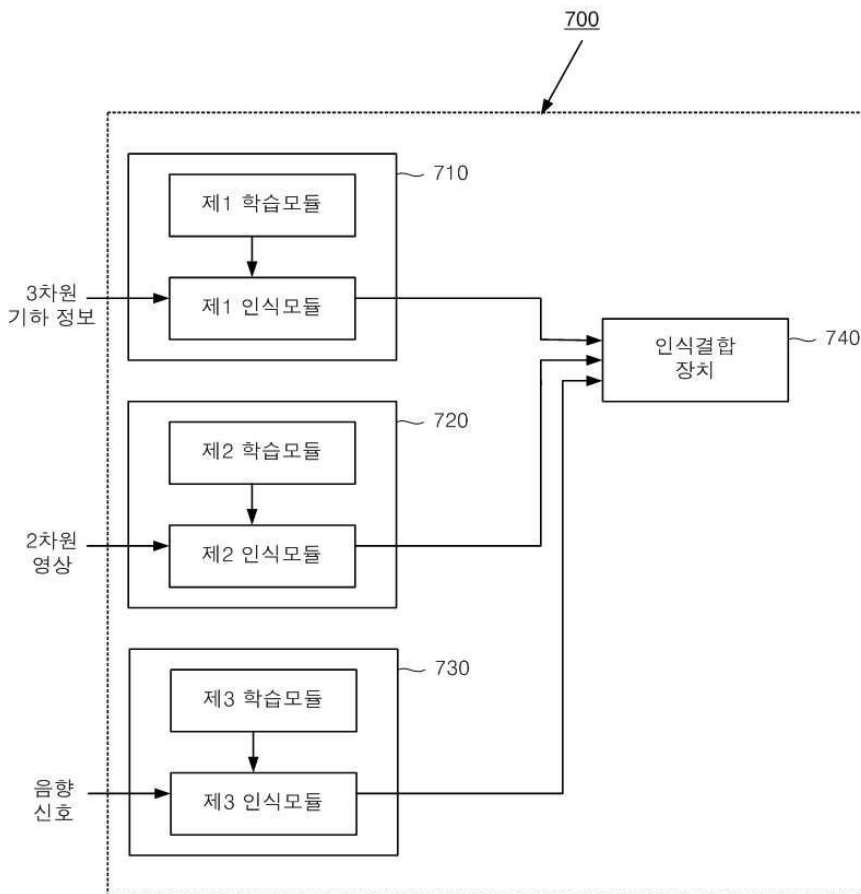
도면9



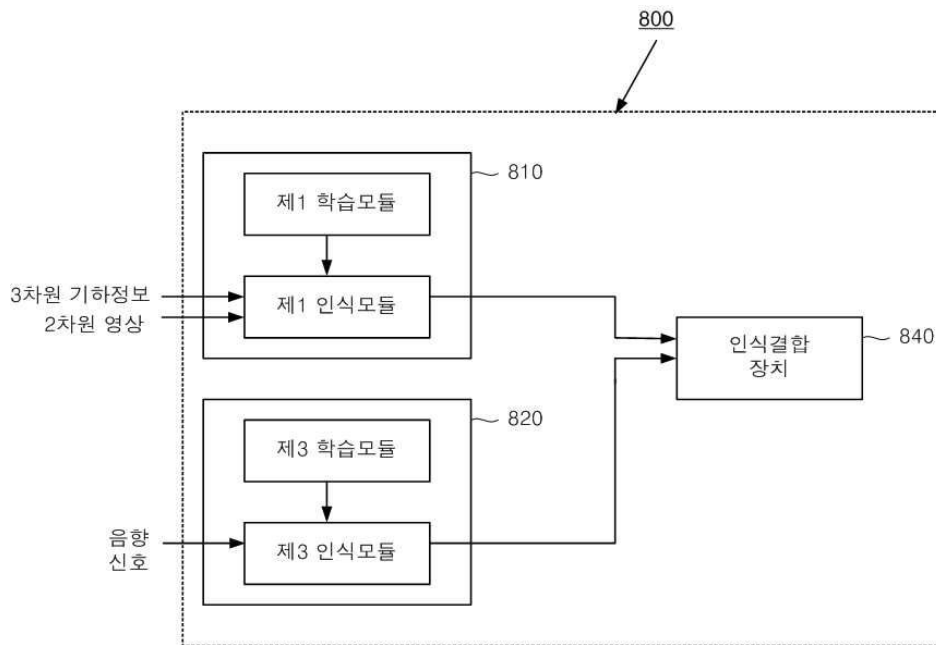
도면10



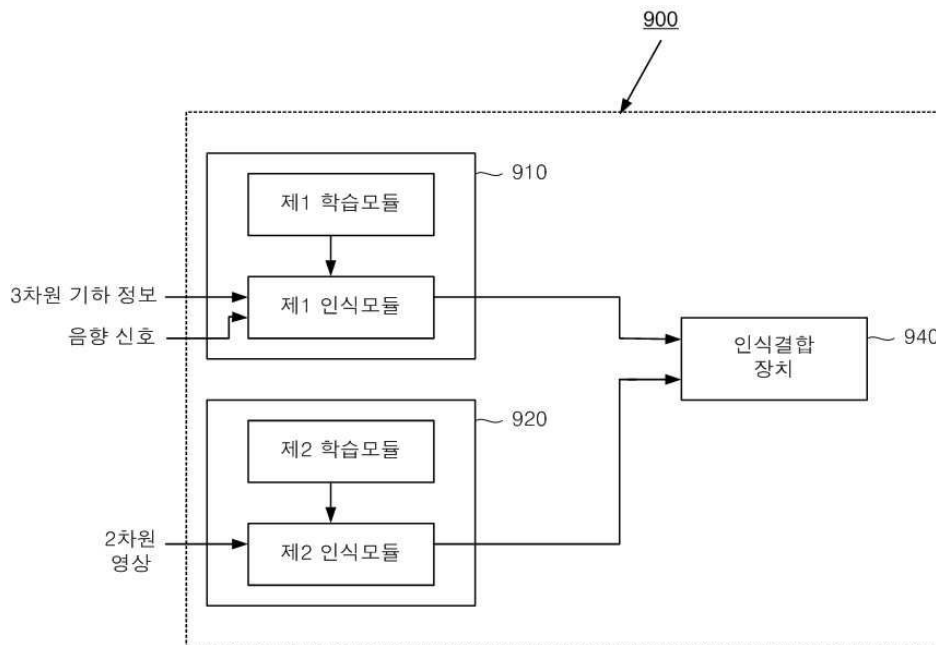
도면11



도면12



도면13



도면14

