



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년11월23일  
(11) 등록번호 10-2181588  
(24) 등록일자 2020년11월16일

(51) 국제특허분류(Int. Cl.)  
G10L 21/16 (2013.01) G06F 3/16 (2018.01)  
G10L 15/07 (2013.01) G10L 15/22 (2006.01)  
G10L 17/24 (2013.01)  
(52) CPC특허분류  
G10L 21/16 (2013.01)  
G06F 3/167 (2013.01)  
(21) 출원번호 10-2019-0157002(분할)  
(22) 출원일자 2019년11월29일  
심사청구일자 2020년02월11일  
(65) 공개번호 10-2019-0135974  
(43) 공개일자 2019년12월09일  
(62) 원출원 특허 10-2017-0055857  
원출원일자 2017년05월01일  
심사청구일자 2017년05월01일  
(30) 우선권주장  
1020160053428 2016년04월29일 대한민국(KR)  
(56) 선행기술조사문헌  
JP2000338995 A\*  
(뒷면에 계속)  
전체 청구항 수 : 총 4 항

(73) 특허권자  
주식회사 브이터치  
서울특별시 강남구 강남대로132길 25, 4층, 5층(논현동, 보인제빌딩)  
(72) 발명자  
김석중  
서울특별시 서초구 서운로 197, 101동 2303호(서초동, 롯데캐슬클래식아파트)  
김정훈  
경기도 성남시 분당구 산운로 97, 503동 1201호(운중동, 한성필하우스아파트)  
김소연  
서울특별시 서초구 서초대로54길 29-5, 402호(서초동)  
(74) 대리인  
모아특허법인

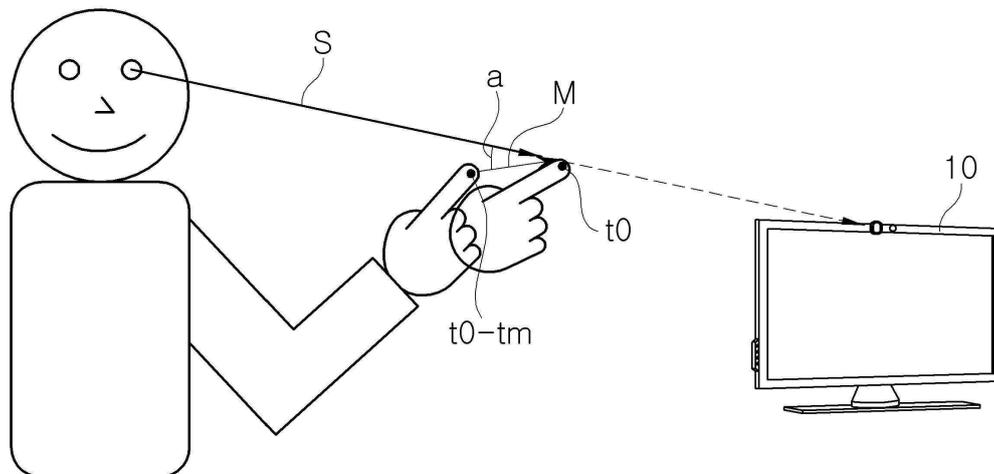
심사관 : 정성윤

(54) 발명의 명칭 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법 및 이를 적용한 전자 장치

(57) 요약

본 발명은 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법 및 이를 적용한 전자 장치에 관한 것으로서, 보다 상세하게는 사용자가 전자 장치를 특정하여 음성 인식 모드로 전환하기 위한 트리거(Trigger)를 동작 인식으로 처리하고, 특정된 전자 장치에 대하여 음성 명령을 전달할 수 있는 제어 방법 및 이를 적용한 전자 장치에 관한 (뒷면에 계속)

대표도 - 도4



한 것이다. 본 발명의 실시예에 따른 최적 제어 방법 및 이를 활용한 전자 장치는 주변의 상황에 상관없이 보다 빠르고 효과적으로 전자 장치에 음성 명령을 전달할 수 있으며, 동작 인식을 통하여 전자 장치를 특정하고 음성 명령을 전달할 수 있도록 함으로써 사용자가 음성 인식을 위해 대상이 되는 전자 장치의 명칭 등을 미리 학습하거나 기억하지 않고도 효과적으로 음성 명령을 수행할 수 있다. 또한, 전자 장치에 음성 명령을 전달하기 위한 사전 단계로서의 동작을 보다 정확하게 인식할 수 있도록 함으로써, 인식률을 높이고 오동작을 방지할 수 있는 효과가 있다.

(52) CPC특허분류

*G10L 15/07* (2013.01)

*G10L 15/22* (2013.01)

*G10L 17/24* (2013.01)

(56) 선행기술조사문헌

KR101453815 B1\*

KR1020120132337 A\*

KR1020140109020 A\*

KR1020150027608 A\*

US20090150160 A1\*

KR1020150065643 A\*

\*는 심사관에 의하여 인용된 문헌

## 명세서

### 청구범위

#### 청구항 1

전자 장치를 제어하는 방법으로서,

사용자의 동작 정보를 참조하여 상기 사용자의 제1 신체 좌표 및 제2 신체 좌표를 획득하는 단계,

상기 제1 신체 좌표로부터 상기 제2 신체 좌표를 잇는 연결 벡터의 크기, 상기 제2 신체 좌표에 기초하여 특정되는 이동 벡터의 크기, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 사용자의 제2 신체의 이동 시간 중 적어도 둘의 조합이 제1 임계 조건을 만족시키는지 여부를 결정하는 단계,

상기 제1 임계 조건을 만족하는 시점으로부터 유효시간 이내에 특정되는 상기 연결 벡터의 크기, 상기 이동 벡터의 크기, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 제2 신체의 이동 시간 중 적어도 둘의 조합이 제2 임계 조건을 만족시키는지 여부를 결정하는 단계, 및

상기 제1 임계 조건 및 상기 제2 임계 조건을 모두 만족시키는 경우에, 상기 사용자의 음성 명령을 인식할 수 있도록 음성 트리거를 생성하는 단계를 포함하고,

상기 제1 임계 조건 또는 상기 제2 임계 조건은, 상기 연결 벡터에 따라 상기 이동 벡터의 특성 값에 대한 범위를 달리하고,

상기 이동 벡터의 특성 값은, 제1 시각과 제2 시각 동안 상기 사용자의 제2 신체 좌표가 이동한 거리, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 사용자의 제2 신체의 포즈 중 적어도 하나를 포함하고,

상기 제1 임계 조건을 만족시키는지 여부를 결정하는 단계에서 특정되는 상기 적어도 둘의 조합과 상기 제2 임계 조건을 만족시키는지 여부를 결정하는 단계에서 특정되는 상기 적어도 둘의 조합은 서로 상이하고,

상기 제1 임계 조건 및 상기 제2 임계 조건은 상기 신체 좌표의 종류에 따라 다르게 설정되는

방법.

#### 청구항 2

제1항에 있어서,

상기 음성 트리거가 생성되면, 상기 사용자의 제2 신체가 움직이는 범위 및 상기 사용자의 제1 신체 좌표에 기초하여 특정되는 명령 범위에 따라, 상기 사용자의 음성 명령이 전달될 영역을 결정하는 단계를 포함하고,

상기 명령 범위는 상기 제2 신체가 움직이는 범위 및 상기 제1 신체 좌표에 기초하여 특정되는 3차원 형상을 포함하는

방법.

#### 청구항 3

전자 장치로서,

사용자의 동작 정보를 참조하여 상기 사용자의 제1 신체 좌표 및 제2 신체 좌표를 획득하는 좌표 검출부, 및

상기 제1 신체 좌표로부터 상기 제2 신체 좌표를 잇는 연결 벡터의 크기, 상기 제2 신체 좌표에 기초하여 특정되는 이동 벡터의 크기, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 사용자의 제2 신체의 이동 시간 중 적어도 둘의 조합이 제1 임계 조건을 만족시키는지 여부를 결정하고, 상기 제1 임계 조건을 만족하는 시점으로부터 유효시간 이내에 특정되는 상기 연결 벡터의 크기, 상기 이동 벡터의 크기, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 제2 신체의 이동 시간 중 적어도 둘의 조합이 제2 임계 조건을 만족시키는지 여부를 결정하고, 상기 제1 임계 조건 및 상기 제2 임계 조건을 모두 만족시키는 경우에, 상기 사용자의 음성 명령을 인식할 수 있도록 음성 트리거를 생성하는 제어부를 포함하고,

상기 제1 임계 조건 또는 상기 제2 임계 조건은, 상기 연결 벡터에 따라 상기 이동 벡터의 특성 값에 대한 범위를 달리하고,

상기 이동 벡터의 특성 값은, 제1 시각과 제2 시각 동안 상기 사용자의 제2 신체 좌표가 이동한 거리, 상기 연결 벡터와 상기 이동 벡터 사이의 각도 및 상기 사용자의 제2 신체의 포즈 중 적어도 하나를 포함하고,

상기 제1 임계 조건을 만족시키는지 여부를 결정하는 단계에서 특정되는 상기 적어도 둘의 조합과 상기 제2 임계 조건을 만족시키는지 여부를 결정하는 단계에서 특정되는 상기 적어도 둘의 조합은 서로 상이하고,

상기 제1 임계 조건 및 상기 제2 임계 조건은 상기 신체 좌표의 종류에 따라 다르게 설정되는 전자 장치.

#### 청구항 4

제3항에 있어서,

상기 음성 트리거가 생성되면, 상기 사용자의 제2 신체가 움직이는 범위 및 상기 사용자의 제1 신체 좌표에 기초하여 특정되는 명령 범위에 따라, 상기 사용자의 음성 명령이 전달될 영역을 결정하고,

상기 명령 범위는 상기 제2 신체가 움직이는 범위 및 상기 제1 신체 좌표에 기초하여 특정되는 3차원 형상을 포함하는

전자 장치.

#### 청구항 5

삭제

#### 청구항 6

삭제

#### 청구항 7

삭제

#### 청구항 8

삭제

#### 청구항 9

삭제

#### 청구항 10

삭제

#### 청구항 11

삭제

#### 청구항 12

삭제

#### 청구항 13

삭제

#### 청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

**발명의 설명**

**기술 분야**

[0001] 본 발명은 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법 및 이를 적용한 전자 장치에 관한 것으로서, 보다 상세하게는 사용자가 전자 장치를 특정하여 음성 인식 모드로 전환하기 위한 트리거(Trigger)를 동작 인식으로 처리하고, 특정된 전자 장치에 대하여 음성 명령을 전달할 수 있는 제어 방법 및 이를 적용한 전자 장치에 관한 것이다.

**배경 기술**

[0002] 최근, 일상 생활에서 사용되고 있는 전자 제품들을 사용자의 음성을 인식하고 이에 따라 동작이 이루어질 수 있도록 하는 시도가 크게 증가되고 있다. 특히, 현대인의 생활에서 가사 노동을 줄일 수 있도록 함으로써 생활의 편리함을 제공하는 세탁기나 식기세척기, 여가 활동이나 정보 수집과 같은 문화나 교육 관련 분야에서 중요한 부분을 차지하고 있는 TV나 오디오(audio) 등에 이러한 음성 인식 기능을 채택하는 경향이 많아지고 있다.

[0003] 이러한 음성 인식은 개략적으로 다음과 같이 동작이 이루어진다. 먼저, 전자 장치에 탑재된 음성 인식 알고리즘은 음성 검출 과정과 특징(feature) 추출 과정, 그리고 매칭(matching) 과정을 순차적으로 거치게 되는데, 마이크를 통해 사용자의 음성 신호가 입력되면 A/D(Analog to Digital) 컨버터에서 이를 디지털 신호로 변환한다. 변환된 디지털 음성 신호는 짧은 구간의 신호(즉, 프레임)로 분할되고, 각 프레임의 에너지와 영 교차율(zero crossing rate) 그리고, 시간 길이 정보 등을 이용하여 입력된 음성 신호 중에서 사용자가 실제로 발성한 음성 구간만을 검출하게 된다. 이후, 검출된 음성 구간에 대한 프레임의 특징을 추출하여 입력된 음성의 테스트 패턴을 만들고, 테스트 패턴과 기준 데이터용 메모리에 저장된 각 기준 패턴들과 각각 비교하여 테스트 패턴과 가장 유사한 특징을 갖는 기준 패턴을 인식된 음성으로 출력하게 되는 것이다. 이 때, 기준 데이터용 메모리에 음성 신호의 기준 패턴을 저장하는 경우에도 음성 구간에 해당하는 프레임의 특징을 추출하여 기준 패턴을 만든 후, 기준 데이터용 메모리에 저장하게 되는데, 이와 같은 동작을 인식하고자 하는 음성 신호에 대하여 반복 수행함으로써 기준 패턴을 기준 데이터용 메모리에 데이터베이스로써 저장하게 되는 것이다.

[0004] 그런데, 음성 인식을 위해서는 사용자의 음성 명령과 일반 음성(대화 등)을 구별하는 과정이 필요하게 된다. 이에 따라, 종래에는 터치 또는 버튼을 누르는 동작으로 음성 인식의 시작 시점을 특정하고, 음성 인식을 통하여 명령을 수행할 전자 장치를 특정한 후에, 음성으로 명령을 입력하여 대상이 되는 전자 장치의 동작을 수행하는 방법을 사용하였다. 등록특허공보 제 10-1000925 호(음성 인식이 효율적으로 이용되는 디지털 촬영 장치의 제어 방법, 및 이 방법을 사용한 디지털 촬영 장치)는 이와 같이, 2단 구조의 셔터 릴리즈 버튼을 구비하고, 셔터 릴리즈 버튼이 눌러지는 상태에 따라 사용자의 음성 명령에 따라 자동 포커싱 또는 촬영을 수행하는 디지털 촬영

장치를 제시하였다.

- [0005] 이와 같이, 음성 인식을 위한 버튼 입력은 일반적으로 홈 네트워크 시스템에서는 리모컨을 통해 가전 기기를 제어할 수 있으며, 자동차와 같은 외부 이동 기기에서는 전자 키를 사용함으로써 동작이 이루어질 수 있다. 그러나, TV와 같은 가전 기기의 경우에는, 무선 리모컨이 수행해야 할 기능들이 점점 더 증가하게 되고, 그에 따라 무선 리모컨이 갖는 버튼의 수도 크게 증가하게 됨으로써, 사용자가 음성 인식 동작을 위해 리모컨을 사용하는 데 불편함을 크게 느끼는 경향이 있다. 특히, 리모컨 기능의 복잡화는 어린이나 노약자에게 더욱 큰 불편함을 가중시킬 수 있으며, 리모컨의 보관 장소를 기억하지 못하거나 분실하는 경우에 리모컨을 이용한 음성 인식 기능의 사용이 어려울 수 있으며, 리모컨의 전원으로 이용되는 건전지가 수명을 다하면 이를 교체해야 하는 등 환경 친화적인 문제점도 있다.
- [0006] 이러한 버튼 입력 방식의 문제점을 개선하기 위하여, 종래에는 음성을 이용하여 사용자가 이용하고자 하는 전자 장치를 특정한 후에, 음성 인식을 통해 명령을 전달하는 방법이 사용되기도 하였다. 이러한 방법은 'Hey Siri', 'Alexa', 'Cortana'...와 같이, 음성 인식을 수행할 전자 장치를 미리 정해진 단어로 등록하고, 사용자가 음성을 통해 전자 장치를 특정하면 특정된 전자 장치가 소리나 LED(Light Emitting Diode) 등을 통해 사용자에게 음성 인식 시작 시점을 지정하게 되고, 그 후에 사용자가 음성으로 명령의 대상이 되는 전자 장치(TV, 오디오, 조명, Fan, Thermostat 등)와 명령어(On, Off, Up, Down 등)를 음성으로 입력하는 방법이다. 하지만 이와 같이, 사용자가 음성을 이용하여 제어하고자 하는 전자 장치를 특정하고 음성 명령을 하는 방법은, 전자 장치의 특정과 음성 입력에 긴 시간과 과정이 소요되는 불편함이 있기 때문에, 음성 인식을 효과적으로 처리하는 방법을 마련하는 것이 필요하다.
- [0007] 특히, 기존의 음성 인식 시스템은 입력된 음성 명령의 주파수를 파악하여 명령을 검색하는데, 비교적 많은 연산량이 요구될 수밖에 없다. 이러한 음성 명령은 주로 전자 장치의 출고 시점에 미리 정해져 있는데, "전원 켜" 또는 "볼륨 업" 등의 간단한 명령어로 구성된다. 그런데, 이와 같이 미리 정해진 음성 명령은 다양한 사용자의 다양한 음색과 발음을 모두 구별하여 정확하게 인식하기 어려운 점이 있으며, 음성 명령이 많은 경우에는 사용자가 수 많은 음성 명령을 모두 암기하고 있어야 하는 불편함이 있다. 또한, 하나의 전자 장치가 다양한 나라로 수출되는 경우, 해당 나라의 언어로 음성 명령을 미리 등록해야 한다는 단점이 있다. 따라서, 음성 명령에 대한 인식 실패율이 높고, 음성을 인식하기 위해 높은 사양의 하드웨어와 복잡한 알고리즘이 요구되는 단점을 극복하기 위한 방안이 필요하다.
- [0008] 이러한 문제점을 해결하기 위해서, 종래에는 사용자의 음성과 동작을 동시에 인식하여 전자 장치를 제어하는 방법이 개시되었다. 등록특허공보 제 10-1046022 호(인체 움직임 및 인간 음성을 이용한 원격 조정 장치)는 사용자의 동작을 감지하여 TV와 같은 디스플레이 장치에 표시되는 마우스 포인터를 제어하는 방법을 개시하였으며, 공개특허공보 제 10-2012-0054743 호 (멀티미디어 장치에서 음성과 제스처를 이용한 제어 방법 및 그에 따른 멀티미디어 장치)는 사용자의 음성과 제스처를 함께 인식하여 음성과 제스처의 특징이 모두 매칭되는 경우에 사용자가 원하는 동작을 실행하는 방법을 개시하였다.
- [0009] 그러나, 일반적으로 동작 인식은 사용자가 공간에서 움직이는 동작을 식별하여야 하는데, 사용자의 동작이 전자 장치의 제어를 위한 동작인지 일반적인 움직임인지를 구별하는 것이 쉽지 않고, 동작 인식을 위해서 사용되는 동작 인식 모듈의 가격도 높기 때문에 전자 장치의 제어를 위하여 일괄적으로 동작 인식 기능을 적용하는 것은 효율성이나 경제적 측면에서 그 효용 가치가 낮은 단점이 있다. 또한, 일반적인 포즈나 동작(신체부위의 이동)은 사용자가 카메라를 향한 상태에서는 상대적으로 잘 인식이 되지만 카메라 및 사람의 위치와 방향이 바뀌면 인식률이 떨어지기 때문에, 홈 네트워크와 같이 여러가지 전자 기기와 사물이 존재하는 공간에서는 각각의 전자 기기 및 사물 마다 복수의 카메라가 구비되어야 하는 단점이 있다.
- [0010] 따라서, 동작 인식과 음성 인식 기능을 병행하는 경우에는 동작 인식으로 처리하는 부분과 음성 인식으로 처리하는 부분을 구분함으로써, 사용자의 이용과 경제적 효율성 등 종합적인 측면을 고려하여 가장 최적화된 처리 프로세스를 구현할 필요가 있다.

**선행기술문헌**

**특허문헌**

- [0011] (특허문헌 0001) 등록특허공보 제 10-1000925 호 (2010.12.07)
- (특허문헌 0002) 등록특허공보 제 10-1046022 호 (2011.06.27)

(특허문헌 0003) 공개특허공보 제 10-2012-0054743 호 (2012.05.31)

**발명의 내용**

**해결하려는 과제**

[0012] 본 발명은 상기와 같은 문제점을 해결하기 위한 것으로서, 사용자가 전자 장치를 특정하여 음성 명령을 전달하기 위한 사전 단계로서의 트리거(Trigger) 이벤트를 동작 인식으로 처리하고, 특정된 전자 장치에 대하여 음성 명령을 전달함으로써, 최적화된 제어 방법 및 이를 활용한 전자 장치를 제공하는데 그 목적이 있다.

**과제의 해결 수단**

[0013] 상기와 같은 목적을 달성하기 위하여, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 동작 인식 장치를 통하여 사용자의 동작 정보를 수신하는 단계와, 상기 동작 정보로부터 사용자의 제1 신체 좌표와 제2 신체 좌표를 검출하는 단계와, 상기 제1 신체 좌표로부터 제2 신체 좌표를 잇는 연결 벡터, 또는 상기 제2 신체 좌표의 이동 벡터를 검출하는 단계와, 상기 이동 벡터의 특성 값이 미리 정해진 임계 조건을 만족하는 경우, 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 전환하는 단계를 포함할 수 있다.

[0014] 상기 사용자의 제1 신체 좌표는 눈, 코 또는 입의 위치 좌표일 수 있다.

[0015] 상기 사용자의 제2 신체 좌표는 사용자의 손가락 끝부분의 위치 좌표일 수 있다.

[0016] 상기 연결 벡터는 상기 제1 신체 좌표에서 상기 제2 신체 좌표까지 이어지는 직선 정보일 수 있다.

[0017] 상기 이동 벡터는 제1 시각과 제2 시각 동안 사용자의 제2 신체 좌표를 연결한 벡터일 수 있다.

[0018] 상기 임계 조건은 상기 연결 벡터에 따라 상기 이동 벡터의 특성 값에 대한 범위를 달리할 수 있다.

[0019] 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 따라, 상기 음성 인식 모드로 전환하는 경우, 특정 시각을 전후하여 사용자의 음성 명령에 대한 인식을 시작할 수 있다.

[0020] 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 상기 제1 신체 좌표와 제2 신체 좌표의 연장 선을 명령 벡터로 설정하고, 상기 명령 벡터의 특성 값이 미리 정해진 실행 조건을 만족하는 경우에, 사용자의 음성 명령을 실행하는 단계들 더 포함할 수 있다.

[0021] 상기 명령 벡터는 이동 벡터의 특성 값이 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계 조건에 만족하면 경우에 검출될 수 있다.

[0022] 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 상기 음성 인식 모드로 전환한 후, 제2 신체 좌표로부터 사용자의 제2 신체가 형성하는 명령 범위에 따라 사용자의 음성 명령을 실행하는 영역을 결정할 수 있다.

[0023] 상기 명령 범위는 사용자의 제1 신체가 향하는 방향에 대하여, 사용자의 제2 신체가 상하좌우로 형성하는 지점과 상기 제1 신체 좌표의 연장 선상이 이루는 공간 영역을 포함할 수 있다.

[0024] 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 상기 음성 명령에 따라 주변에 위치하는 전자 장치의 동작을 제어하는 단계를 더 포함할 수 있다.

[0025] 상기 음성 명령에 따라 주변에 위치하는 전자 장치의 동작을 제어하는 단계는 상기 이동 벡터의 특성 값이 미리 정해진 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계조건에 만족하는 경우, 사용자의 음성 명령을 인식할 수 있도록 주변에 위치하는 전자 장치를 음성 인식 모드로 전환시키는 단계를 포함할 수 있다.

[0026] 또한, 본 발명의 동작-음성의 다중 모드 명령에 기반한 전자 장치는 사용자의 동작 정보를 인식하기 위한 동작 인식부와, 사용자의 음성 명령을 인식하기 위한 음성 인식부와, 상기 동작 인식부로부터 인식된 동작 정보로부터 사용자의 제1 신체 좌표와 제2 신체 좌표를 검출하는 좌표 검출부와, 상기 제1 신체 좌표로부터 제2 신체 좌표를 잇는 연결 벡터, 또는 상기 제2 신체 좌표의 이동 벡터를 검출하는 벡터 검출부와, 상기 이동 벡터의 특성 값이 미리 정해진 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계조건에 만족하는 경우, 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 전환하는 제어부를 포함할 수 있다.

- [0027] 상기 동작 인식부는 카메라 모듈 및, 동작 센서로 이루어질 수 있다.
- [0028] 상기 음성 인식부는 마이크 및 음성 센서를 포함할 수 있다.
- [0029] 또한, 본 발명의 동작-음성의 다중 모드 명령에 기반한 전자 제어 장치는 상기 음성 인식부를 통해 입력된 사용자의 음성 명령에 따라 주변에 위치하는 전자 장치의 동작을 제어하는 제어부를 더 포함할 수 있다.
- [0030] 또한, 본 발명의 동작-음성의 다중 모드 명령에 기반한 전자 제어 장치는 상기 이동 벡터의 특성 값이 미리 정해진 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계조건에 만족하는 경우, 사용자의 음성 명령을 인식할 수 있도록 주변에 위치하는 전자 장치를 음성 인식 모드로 전환시킬 수 있다.

**발명의 효과**

- [0031] 본 발명의 실시예에 따른 최적 제어 방법 및 이를 활용한 전자 장치는 주변의 상황에 상관없이 보다 빠르고 효과적으로 전자 장치에 음성 명령을 전달할 수 있으며, 동작 인식을 통하여 전자 장치를 특정하고 음성 명령을 전달할 수 있도록 함으로써 사용자가 음성 인식을 위해 대상이 되는 전자 장치의 명칭 등을 미리 학습하거나 기억하지 않고도 효과적으로 음성 명령을 수행할 수 있다. 또한, 전자 장치에 음성 명령을 전달하기 위한 사전 단계로서의 동작을 보다 정확하게 인식할 수 있도록 함으로써, 인식률을 높이고 오동작을 방지할 수 있는 효과가 있다.

**도면의 간단한 설명**

- [0032] 도 1은 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 전체적인 개념도,
- 도 2는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 적용된 전자 장치의 내부 구성도,
- 도 3은 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 가진 제품에 적용된 경우의 예시도,
- 도 4는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도,
- 도 5는 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도,
- 도 6은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도,
- 도 7은 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 전자 장치의 음성 인식 기능을 트리거시키기 위한 조건을 예시로 나타낸 도표,
- 도 8은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 음성 인식을 위한 트리거 이후에 명령 벡터를 이용하여 전자 장치를 제어하는 경우의 설명도,
- 도 9 내지 도 11은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 사용자가 일정한 영역 내에 있는 전자 장치에 대해 음성 명령을 내리는 경우를 설명하기 위한 도면,
- 도 12는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 흐름도,
- 도 13은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 흐름도,
- 도 14는 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 흐름도,
- 도 15는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 활용한 전자 장치의 소프트웨어 구성 예시도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0033] 여기서 사용되는 전문용어는 단지 특정 실시예를 언급하기 위한 것이며, 본 발명을 한정하는 것을 의도하지 않는다. 여기서 사용되는 단수 형태들은 문구들이 이와 명백히 반대의 의미를 나타내지 않는 한 복수 형태들도 포함한다. 명세서에서 사용되는 "포함하는"의 의미는 특정 특성, 영역, 정수, 단계, 동작, 요소 및/또는 성분을

구체화하며, 다른 특정 특성, 영역, 정수, 단계, 동작, 요소, 성분 및/또는 군의 존재나 부가를 제외시키는 것은 아니다.

- [0034] 다르게 정의하지는 않았지만, 여기에 사용되는 기술용어 및 과학용어를 포함하는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 일반적으로 이해하는 의미와 동일한 의미를 가진다. 보통 사용되는 사전에 정의된 용어들은 관련 기술문헌과 현재 게시된 내용에 부합하는 의미를 가지는 것으로 추가 해석되고, 정의되지 않는 한 이상적이거나 매우 공식적인 의미로 해석되지 않는다.
- [0035] 이하, 첨부한 도면에 의하여 본 발명의 바람직한 실시예를 자세히 설명하도록 한다.
- [0037] 도 1은 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 전체적인 개념도이다. 도 1을 참조하면, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 다양한 가정용 기기를 제어하는데 적용될 수 있는데, 대표적인 적용대상이 홈 네트워크 시스템이 될 수 있다.
- [0038] 즉, 다수의 가정용 기기를 포함하는 홈 네트워크 시스템의 일정 영역에 본 발명의 전자 장치(100)가 배치될 수 있다. 홈 네트워크 시스템은 유선 또는 무선 네트워크를 통해 연결되는 복수의 기기를 포함할 수 있다. 네트워크는 가정 내에서 사용되는 모든 전기, 전자 기기들이 유선 또는 무선으로 연결되어 양방향 통신이 가능하도록 설정되는 네트워크이다. 이러한 네트워크는 유선 또는 무선 인터넷을 사용할 수 있으며, 이를 통해 TV(10), 에어컨(20), 조리 기기(30), 노트북(40), 로봇 청소기(50), 식기 세척기(60), 냉장고(70), 전자레인지(80), 또는 세탁기(90) 등의 전자 장치 사이에 정보 공유가 가능하도록 한다. 유선 네트워크의 대표적인 것으로 기존 전화를 이용한 홈 PNA(Home Phoneline Networking Alliance), IEEE 1394, 전력선 방식을 들 수 있고, 무선 네트워크의 형태로는 홈 RF(Home Radio Frequency), 블루투스(Bluetooth), IrDA(Infrared Data Association) 및 무선 LAN(Wifi) 등을 들 수 있다. 사용자는 본 발명의 전자 장치(100)를 가정 내에 배치하고, 홈 네트워크 시스템을 구성하는 가정용 기기를 동작과 음성으로 제어할 수 있다. 또한, 본 발명의 전자 장치(100)는 별도의 기기로 구현될 수도 있지만, 가정용 기기 등 기존의 기기 내에 그 기능을 내장하여 구현할 수도 있다.
- [0039] 도 1에서는 홈 네트워크 시스템의 경우를 예로 들어 설명하였지만, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 가정 내에서 뿐만 아니라, 스마트 자동차, 드론, 디지털 카메라 등 가정 외부에서 사용되는 전자 장치에도 적용될 수 있는 것은 자명할 것이다.
- [0041] 도 2는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 적용된 전자 장치의 내부 구성도를 나타낸 것이다. 도 2를 참조하면, 본 발명의 전자 장치(100)는 메모리 시스템(150)과, 여기에 연결되어 전자 장치(100)의 동작을 제어하는 제어부(140), 입력 장치(110) 및 출력 장치(120)를 포함한다.
- [0042] 여기에서, 입력 장치(110)는 사용자의 움직임에 따라 영상 정보를 입력받기 위한 카메라(112)와 사용자의 음성 정보를 입력받기 위한 마이크(114)를 포함하고, 추가적으로 터치 스크린과 같이 사용자의 터치를 인식할 수 있는 장치 또는, 통상적인 마우스나 전자 펜과 같은 여러가지 형태의 물리적 변환기(Physical transducer) 등도 포함될 수 있다. 출력 장치(120)는 스마트폰이나 태블릿의 경우에 문자나 이미지를 표시할 수 있는 디스플레이를 의미하지만, 프린터나 스피커와 같은 변환기(transducer)도 포함될 수 있다. 디스플레이는 액정 디스플레이(Liquid Crystal Display, LCD), 박막 트랜지스터 액정 디스플레이(Thin Film Transistor-Liquid Crystal Display, TFT LCD), 유기 발광 다이오드(Organic Light-Emitting Diode, OLED), 플렉서블 디스플레이(flexible display), 3차원 디스플레이(3D display) 등 다양한 형태로 이루어질 수 있다. 전자 장치(100)가 통신 장치인 경우에는 네트워크 인터페이스 또는 모뎀과 같은 장치가 입력 및/또는 출력 장치로서 사용될 수도 있다.
- [0043] 입력 장치(110)에는 센서부(130)가 연결되는데, 센서부(130)는 입력 장치(110)를 통해 사용자의 움직임을 인식하기 위한 동작 센서(132)와 사용자의 음성 신호를 인식하기 위한 음성 센서(134)를 함께 구비한다. 입력 장치(110)를 통해 사용자의 움직임이 인식되거나 사용자의 음성 신호가 인식되는 경우, 센서부(130)는 동작 정보 및 음성 정보를 제어부(140)에 제공하게 된다.
- [0044] 제어부(140)는 전자 장치(100)의 작동을 제어하기 위한 CPU(Central Processing Unit; 142)와, 카메라(112) 및 동작 센서(132)를 통해 인식된 사용자의 동작 방향 및 크기를 산출하는 영상 신호 처리부(144)와, 음성 센서(134)를 통해 전달되는 사용자의 디지털 음성 신호로부터 음성 패턴을 산출하는 음성 신호 처리부(146)를 포함한다. 영상 신호 처리부(12)는 카메라(112)로부터 입력되는 사용자의 영상 신호에서 움직임이 발생하는 영역을 식별하고, 이 움직임이 발생하는 영역의 방향과 크기를 이동 벡터로 산출한다. 바람직하게, 영상 신호 처리부(144)는 산출된 이동 벡터의 크기가 미리 설정된 기준 값보다 작은 경우에는 전자 장치(100)의 원격 조정을 위한 동작이 아닌 것으로 판단하고 무시할 수 있으며, 산출된 이동 벡터의 크기가 미리 설정된 기준 값보다 큰 경

우에는 사용자가 전자 장치(100)의 원격 조정을 위한 동작인 것으로 판단하고 지정된 명령을 수행할 수 있다. 일반적으로, 마이크(114)를 통해 입력되는 음성 신호는 아날로그 신호이기 때문에, 음성 센서(134) 및 음성 신호 처리부(146)에 입력될 수 있도록 아날로그/디지털 변환기(ADC)(도면에 도시되지 않음)가 구비될 수 있다. 음성 신호 처리부(146)는 입력된 음성 신호로부터 음성 패턴을 산출한 후, 메모리 시스템(150)에 미리 저장되어 있는 음성 패턴과 비교하여, 일치하는 음성 패턴이 존재하는 경우에 해당 음성 패턴에 대응되는 명령을 수행하도록 한다.

[0045] 또한, 제어부(140)는 데이터 및 명령어의 일시적인 저장을 위한 레지스터와 전자 장치(100)의 동작을 제어하기 위한 컨트롤러를 더 포함할 수 있다. 한편, CPU(142)는 디지털(Digital) 사의 알파(Alpha), MIPS 테크놀로지, NEC, IDT, 지멘스(Siemens) 등의 MIPS, 인텔(Intel)과 사이릭스(Cyrix), AMD 및 넥스젠(Nexgen)을 포함하는 회사의 x86 및 IBM과 모토롤라(Motorola)의 파워PC(PowerPC)와 같이 다양한 아키텍처(Architecture)를 갖는 프로세서일 수 있다.

[0046] 메모리 시스템(150)은 일반적으로 RAM(Random Access Memory) 과 ROM(Read Only Memory), SRAM(Static Random Access Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory) 같은 저장 매체 형태인 고속의 메인 메모리(152)와, 플로피 디스크, 하드 디스크, 테이프, CD-ROM, 플래시 메모리, 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리) 등의 장기(longterm) 저장 매체 형태의 보조 메모리(154) 및 전기, 자기, 광학이나 그 밖의 저장 매체를 이용하여 데이터를 저장하는 장치를 포함한다. 또한, 메인 메모리(152)는 디스플레이 장치를 통하여 이미지를 디스플레이 하는 비디오 디스플레이 메모리를 포함할 수 있다. 본 발명의 기술 분야에서 통상의 지식을 가진 당업자에게는 상기 메모리 시스템(150)이 여러 가지 저장 성능을 구비하는 제품으로서, 여러 가지 형태를 가질 수 있다는 것이 자명할 것이다.

[0047] 따라서, 사용자가 입력 장치(110)를 통해 전달하는 동작 또는 음성 신호는 센서부(130)를 통해 제어부(140)에 전달되고, 제어부(140)는 이로부터 산출된 이동 벡터 정보를 메모리 시스템(150)에 기 저장된 동작 및 음성 신호의 기준 값과 비교하여, 일치 여부를 판단하고 일치하는 경우에 원격 제어 권한을 가진 사용자로 판단하여 지정된 명령을 수행하게 될 것이다.

[0048] 본 발명의 기술 분야에 있어서, 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 적용된 전자 장치(100)는 OS(Operating System) 및 적어도 하나의 응용 프로그램을 포함할 수 있다. OS는 전자 장치(100)의 동작 및 리소스의 지정을 제어하는 소프트웨어 집합이다. 응용 프로그램은 OS를 통하여 이용 가능한 컴퓨터 리소스를 사용함으로써, 사용자가 요청한 업무를 수행하기 위한 소프트웨어 집합이다. OS 및 응용 프로그램은 메모리 시스템(150)에 상주될 것이다. 컴퓨터 프로그래밍의 기술 분야에서 통상의 지식을 가진 당업자의 경험에 따라, 다른 표현으로 기술되지 않으면 본 발명은 전자 장치(100)에 의해 수행되는 동작 및 동작에 대한 표현 기호에 따라 기술될 것이다. 이러한 동작은 컴퓨터 기반으로 이루어지며, OS 또는 적당한 응용 프로그램에 의하여 수행될 것이다. 또한, 이러한 동작 및 기능은 전기 신호의 변환 또는 차단을 유발하는 데이터 비트 등의 전기 신호에 대한 제어부(140)의 처리와, 전자 장치(100)의 동작을 변경할 뿐만 아니라 메모리 시스템(150) 내의 메모리 영역에 저장된 데이터 비트 신호에 대한 관리를 포함한다. 데이터 비트 신호가 관리되는 메모리 영역은 데이터 비트에 해당하는 전기, 자기 또는 광학 특성을 갖는 물리 영역이다.

[0050] 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법은 홈 네트워크에 구비되는 가전 제품이나 그 밖의 다양한 전자 제품내에 구현될 수도 있고, 가정 내에 별도로 설치하여 주변의 전자 제품을 제어하기 위한 독립적인 전자 장치 내에 구현될 수도 있다. 만약, 주변의 전자 제품에 대한 음성 명령을 제어하기 위한 별도의 독립적인 전자 장치로 구현되는 경우에는, 일정한 거리 내에 있는 전자 제품의 동작을 제어하기 위한 무선 통신 기능과 주변에 있는 전자 제품을 식별할 수 있는 기능을 포함할 수 있을 것이다.

[0052] 도 3은 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 가전 제품에 적용된 경우의 예시도이다. 도 3을 참조하면, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 적용된 전자 장치(100)는 TV와 같은 가정용 기기에 구현될 수도 있고, 가정용 기기를 제어하기 위한 별도의 제어 모듈로 구현될 수도 있다.

[0053] 본 발명의 전자 장치(100)는 사용자의 동작을 인식하기 위한 카메라(112)와 사용자의 음성을 인식하기 위한 마이크(114)를 구비할 수 있다. 통상적으로 카메라(112)는 입력되는 빛을 집광하는 렌즈와 렌즈에 의해 초점이 형성되는 영상을 검출하여 전기적인 이미지 신호로 변환하는 CCD(Charge Coupled Device) 또는 CMOS(Complementary Metal-Oxide Semiconductor)와 같은 센서를 포함할 수 있다. 마이크(114)는 사용자로부터

입력되는 아날로그 음성 신호를 디지털 음성 신호로 변환하는 아날로그/디지털 변환기(ADC)를 포함할 수 있다. 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법이 적용된 전자 장치(100)에 카메라(112)와 마이크(114)를 부착하는 경우, 사용자의 동작이나 음성에 대한 인식이 잘 이루어질 수 있도록 TV 프레임의 상단에 설치될 수 있다. 이에 따라, 카메라(112)는 전자 장치(100)의 전방에서 전자 장치(100)를 원격으로 제어하고자 하는 사용자의 움직임을 보다 쉽게 검출할 수 있을 것이다. 사용자가 움직이는 동작은 여러 가지 방법이 있을 수 있지만, 원격 제어의 대상이 되는 전자 장치(100)를 지정하는 사용자의 손가락 움직임이 될 수 있다. 사람의 손가락은 간단한 움직임만으로도 충분한 크기와 방향성을 제공할 수 있기 때문에, 본 발명의 전자 장치(100)에 설치되는 카메라(112)에 의해 검출되는 신체의 움직임은 사용자의 손가락 움직임을 대상으로 하는 것이 바람직하다.

[0055] 도 4는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도이다. 도 4를 참조하면, 사용자는 원격 위치에 있는 특정 기기(10)를 대상으로, 손가락을 이용한 동작을 통해 대상이 되는 기기(10)의 동작을 제어할 수 있으며, 그 과정을 설명하면 아래와 같다. 전술한 바와 같이, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 기능은 일반 기기에 탑재될 수도 있고, 일반 기기를 제어하기 위한 별도의 제품으로 구성될 수도 있다.

[0056] 먼저, 사용자의 한쪽 눈에서 사용자의 손가락까지의 직선이 연결 벡터(S)가 된다. 이 상태에서, 사용자가 손가락을 움직이면 이동 직선에 해당하는 이동 벡터(M)가 생성된다. 이 상태에서, 사용자가 손가락을 이동하여 연결 벡터(S)까지 움직이면 이동 직선에 해당하는 이동 벡터(M)가 생성된다. 이 과정에서, 연결 벡터(S)의 크기, 사용자가 손가락을 움직인 시간( $t_m$ )과 이동 벡터(M)의 크기 및 연결 벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도( $a$ )를 계산할 수 있다. 이 과정에서, 연결 벡터(S)의 크기, 사용자가 손가락을 움직인 시간( $t_m$ )과 이동 벡터(M)의 크기, 연결 벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도( $a$ )의 값이 본 발명의 전자 장치(100)에 저장된 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계조건에 만족하는 경우 본 발명의 전자 기기(100)의 음성 인식 기능을 트리거시키게 된다. 이 때, 기준이 되는 신체 좌표는 사용자의 신체 특징 중 하나를 선택할 수 있으며, 연결 벡터(S)의 크기, 손가락을 움직인 시간( $t_m$ ), 이동 벡터(M)의 크기, 연결 벡터(S)와 이동 벡터(M)가 형성하는 각도( $a$ )의 기준 값은 신체 좌표의 종류에 따라 다르게 설정될 수 있을 것이다.

[0058] 도 5는 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도이다. 도 5를 참조하면, 사용자의 입에서 사용자의 손가락까지의 직선이 연결 벡터(S)가 된다. 이 상태에서, 사용자가 손가락을 움직이면 이동 직선에 해당하는 이동 벡터(M)가 생성된다. 이 과정에서, 연결 벡터(S)의 크기, 사용자가 손가락을 움직인 시간( $t_m$ )과 이동 벡터(M)의 크기 및 연결 벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도( $a$ )를 계산할 수 있으며, 계산된 값들이 본 발명의 전자 장치(100)에 저장된 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계조건에 만족하는 경우 상기 전자 장치(100)의 음성 인식 기능을 트리거시키게 된다.

[0060] 도 6은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 설명하기 위한 예시도이다. 도 6을 참조하면, 사용자의 코에서 사용자의 손가락까지의 직선이 연결 벡터(S)가 된다. 이 상태에서, 사용자가 손가락을 움직이면 이동 직선에 해당하는 이동 벡터(M)가 생성된다. 이에 따라 연결 벡터(S)와 이동 벡터(M)의 크기, 사용자가 손가락을 움직인 시간( $t_m$ ), 연결 벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도( $a$ ) 및 기준 시점( $t_0$ )에 대한 결정은 도 4와 도 5에서 설명한 프로세스가 동일하게 적용된다.

[0062] 도 7을 참조하면, 사용자가 한쪽 눈을 제1 신체 좌표로 하고 손끝을 제2 신체 좌표로 설정하는 경우, 손의 포즈(HP)가 검지 손가락 하나를 편 상태이고, 손끝점이 0.1초 동안 움직인 이동 벡터(M)의 크기가 3cm 이상이고, 이동 벡터(M)와 연결 벡터(S)가 공간 상에서 이루는 각도가 -40도~40도 사이의 값을 가지면 제1 임계 조건을 만족하는 것으로 본다.

[0063] 해당 시점으로부터 0.2초의 유효시간 이내에 손의 포즈(HP)가 검지 손가락 하나를 편 상태이고, 손끝점이 0.1초 동안 움직인 이동 벡터(M)의 크기가 1cm 미만이고, 연결 벡터(S)의 크기가 10~100cm 사이의 값을 가지면 제2 임계 조건을 만족하는 것으로 보고 제1 명령벡터를 검출하는 것으로 본다.

[0064] 사용자가 코 또는 입을 제1 신체 좌표로 하고 손끝을 제2 신체 좌표로 설정하는 경우, 손의 포즈(HP)가 검지 손가락 하나를 편 상태이고, 손끝점이 0.1초 동안 움직인 이동 벡터(M)의 크기가 3cm 이상이고, 이동 벡터(M)와 연결 벡터(S)가 공간 상에서 이루는 각도가 140도~220도 사이의 값을 가지면 제1 임계 조건을 만족하는 것으로 본다.

- [0065] 해당 시점으로부터 0.2초의 유효시간 이내에 손의 포즈(HP)가 검지 손가락 하나를 편 상태이고, 손끝점이 0.1초 동안 움직인 이동 벡터(M)의 크기가 1cm 미만이고, 연결 벡터(S)의 크기가 5cm 미만의 값을 가지면 제2 임계 조건을 만족하는 것으로 보고 제2 명령벡터를 검출하는 것으로 본다.
- [0066] 이러한 조건은 다양한 값으로 변경하여 설정할 수 있는 것은 자명할 것이다.
- [0067] 위와 같이, 사용자의 신체 좌표를 기준으로 손가락까지의 연결 벡터(S), 손가락의 이동 벡터(M)에 대한 크기, 이동 시간 및 연결 벡터(S)와 이동 벡터(M) 사이의 각도가 제1 임계 조건을 만족하고 유효시간 이내에 제2 임계 조건에 만족하는 경우, 전자 장치(100)는 사용자의 음성을 인식하기 위한 모드로 변경하고, 사용자가 음성 신호를 입력하면 이를 인식하여 인식된 결과 값에 해당하는 명령을 실행하게 된다.
- [0068] 따라서, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 따르면, 사용자는 주위의 소음이나 잡음이 있는 경우에도 특정 전자 장치의 음성 인식 기능을 간단한 동작으로 트리거시킬 수 있게 되어, 효과적인 음성 명령을 실행할 수 있다. 또한, 본 발명의 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 활용한 전자 장치는 음성 인식을 위한 트리거 수단으로 사용자의 동작을 활용하고, 음성 인식 기능으로 트리거 된 이후에는 음성을 활용함으로써, 사용자가 음성 인식을 위해 대상이 되는 전자 장치의 명칭 등을 미리 학습하거나 기억하지 않고도 효과적으로 음성 명령을 수행할 수 있으며, 전자 장치의 인식률을 높이고 오동작을 방지할 수 있다.
- [0070] 도 8은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 음성 인식을 위한 트리거 이후에 명령 벡터를 이용하여 전자 장치를 제어하는 경우의 설명도이다. 도 8은, 사용자도 5 내지 도 7의 과정 중 어느 하나의 과정을 통하여, 특정한 전자 장치가 음성 인식을 위한 모드로 트리거시킨 경우를 나타낸 것이다. 따라서, 전자 장치는 사용자의 음성 인식을 위한 모드로 이미 전환된 상태이다. 이때, 사용자는 별도의 동작이 없이도 음성으로 명령을 내릴 수 있겠지만, 주변에 여러 대의 전자 장치가 있는 경우에는 하나의 전자 장치만을 특정해서 음성 명령을 내리고자 할 수가 있다. 이를 위해서는 음성 인식 모드로 전환된 전자 장치를 대상으로 사용자가 신체 좌표(여기서는 한쪽 눈)와 손가락을 잇는 연장선을 명령 벡터(c)로 설정하고, 명령 벡터(C)의 연장선 상에 있는 전자 장치만이 사용자의 음성 명령을 수행하도록 하는 것이다.
- [0072] 도 8은 여러 대의 전자 장치 중에서 하나의 전자 장치만을 특정해서 사용자가 음성 명령을 내리는 경우를 설명하였지만, 반대로 사용자가 일정한 영역에 있는 전자 장치를 대상으로 한꺼번에 음성 명령을 내리는 경우도 가능할 것이다.
- [0073] 도 9 내지 도 11은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 사용자가 본 발명의 전자 장치(100)를 이용하여 일정한 영역 내에 있는 가정용 기기에 대해 음성 명령을 내리는 경우를 설명하기 위한 도면이다. 도 9 내지 도 11을 참조하면, 사용자가 코를 신체 좌표로 설정한 경우를 가정한다. 사용자는 가정용 기기를 음성 인식 모드로 트리거시키기 위한 기준 시점(t0) 또는 음성 인식 모드로 트리거된 후 음성 명령을 내리기 위한 명령 시점(tc)에 손가락을 전자 장치가 있는 상하 방향 또는 좌우 방향으로 설정할 수 있다. 이와 같이, 사용자가 움직인 손가락의 범위는 사용자의 신체 좌표(코)로부터 연장되어 가상의 평면까지 이어지면 전체적으로 원뿔 모양의 음성 명령 범위(Field of Command, FoC)가 형성된다. 이때, 형성되는 원뿔의 면적은 신체 좌표로부터 가상의 평면까지의 거리(r), 평면의 상하 높이(h), 및 평면의 좌우 너비(w)로 결정될 수 있으며, 이러한 공간에 위치하는 전자 장치는 사용자의 음성 명령을 일괄적으로 수행하도록 설정될 수 있다. 도 10의 경우에는 음성 명령의 범위에 위치하는 TV가 사용자의 음성 명령을 수행할 것이며, 음성 명령의 범위에서 벗어나 있는 로봇 청소기는 사용자의 음성 명령을 수행하지 않고 무시하게 될 것이다.
- [0074] 한편, 도 10은 사용자가 음성 명령의 범위를 피라미드 형태로 설정하는 경우를 나타낸 것이며, 도 11은 음성 명령의 범위를 타원의 형태로 설정하는 경우를 나타낸 것이다. 도 10의 경우에는 로봇 청소기가 음성 명령의 범위에 위치하지 않고 있지만, 도 11의 경우에는 로봇 청소기가 음성 명령의 범위에 위치하고 있어서 사용자의 음성 명령을 수행하게 될 것이다.
- [0076] 도 12는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법의 흐름도를 나타낸 것이다. 도 12를 참조하면, 전자 장치(100)는 부착된 카메라(112)를 통하여 주변에 있는 사용자의 영상을 획득한다. 이때, 기준이 되는 사용자의 신체 좌표(눈, 코 또는 입 등)가 영상에 포착되는 경우에는 신체 좌표를 산출하고, 신체 좌표를 기준으로 사용자의 손가락까지 이어지는 연결 벡터(S)와 사용자의 손가락이 이동하는 이동 벡터(M)를 산출한다. 이때, 연결 벡터(S)와 이동 벡터(M)의 크기, 사용자가 손가락을 움직인 시간(tm), 연결

벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도(a) 및 기준 시점(t0)이 메모리 시스템(150)에 저장된 임계 조건의 범위에 해당하면, 사용자가 음성 명령을 위한 트리거를 요청한 것으로 판단하여 음성 명령을 인식할 수 있도록 음성 트리거를 생성한다. 음성 트리거에 의해 음성 인식 모드로 전환된 후 사용자의 음성 명령이 있으면, 부착된 마이크(114)를 통하여 사용자의 음성을 인식하고 음성 인식 결과에 따라 음성 명령을 실행하게 된다. 이 때, 특정 전자 장치에 대하여 음성 명령을 전달하고자 하는 경우, 주변의 다른 전자 장치에 의한 소음이나 잡음이 방해가 될 수 있으므로, 음성 명령 전달 대상이 되는 전자 장치 이외의 전자 장치에 대해서는 음향을 소거하거나 감소시키는 명령을 생성하는 것도 가능할 것이다.

[0078] 도 13은 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 음성 인식을 위한 트리거 이후에 명령 벡터를 이용하여 전자 장치를 제어하는 경우의 흐름도를 나타낸 것이다. 도 13을 참조하면, 사용자의 신체 좌표를 기준으로 손가락의 동작을 인식하여 음성 명령을 위한 임계 조건의 범위에 해당하는 경우에는 사용자의 음성을 인식할 수 있도록 음성 트리거가 생성된다.

[0079] 음성 트리거를 통해 음성 인식 모드로 전환한 시점부터 일정 시간이 흐른 음성 명령 시작 지점(tc)에서 사용자의 음성 명령 뿐만 아니라 사용자의 신체 좌표(눈, 코, 또는 입)와 손가락의 동작을 반영한 명령 벡터(c)를 함께 고려하게 된다. 이에 따라, 명령 벡터(C)의 연장선 상에 있는 전자 장치만이 사용자의 음성 명령을 수행하게 되는 것이다. 명령 벡터(c)는 앞에서 설명한 바와 같이, 전자 장치가 음성 인식 모드로 전환된 상태에서 사용자가 손가락을 이동한 시간(tm), 손가락의 이동 거리 등을 판단 조건으로 할 수 있을 것이다.

[0081] 도 14는 본 발명의 또 다른 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법에 있어서, 사용자가 일정한 영역 내에 있는 전자 장치에 대해 음성 명령을 내리는 경우의 흐름도이다.

[0082] 사용자의 신체 좌표를 기준으로 손가락의 움직임에 판단하여, 신체 좌표와 손가락 사이의 연결 벡터(S)와 사용자의 손가락이 이동하는 이동 벡터(M)가 산출된다. 이 때, 연결 벡터(S)와 이동 벡터(M)의 크기, 사용자가 손가락을 움직인 시간(tm), 연결 벡터(S)와 이동 벡터(M)가 공간상에서 형성하는 각도(a) 및 기준 시점(t0)이 메모리 시스템(150)에 저장된 임계 조건의 범위에 해당하면, 사용자의 음성 명령을 인식할 수 있도록 음성 트리거를 생성한다.

[0083] 음성 트리거에 의해 음성 인식 모드로 전환되면 사용자가 음성 명령을 내릴 범위를 설정하는지를 판단하게 되는데, 앞에서 설명한 바와 같이 사용자의 손가락이 공간상에 그리는 영역을 기준으로 명령 범위를 결정하게 된다. 사용자가 설정한 명령 범위에 위치하는 전자 장치는 사용자의 음성 명령을 인식하여 음성 명령에 따른 동작을 수행하게 되고, 명령 범위에 위치하지 않는 전자 장치는 사용자의 음성 명령을 무시하게 될 것이다. 만약, 사용자의 동작이 명령 범위를 설정하는 조건에 만족하지 않는 경우에는 영역을 고려하지 않고, 사용자가 손가락을 이용하여 명령 벡터(c)를 지정한 전자 장치만이 음성 명령을 수행하게 될 것이다.

[0085] 도 15는 본 발명의 실시예에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 활용한 전자 장치의 소프트웨어 구성 예시도이다. 본 발명의 일례에 따른 동작-음성의 다중 모드 명령에 기반한 최적 제어 방법을 활용한 전자 장치의 응용 모듈(380)은 사용자가 신체 좌표를 기준으로 손가락을 움직이는 동작을 인식하는 동작 인식 모듈(381)과, 음성 인식 모드로 전환 후 사용자가 마이크를 통해 전달하는 음성 명령을 입력받아 이를 처리하는 음성 인식 모듈(382) 및 사용자가 입력한 음성 명령에 따라 전자 장치의 동작을 제어하는 명령 처리 모듈(383)을 포함할 수 있다.

[0086] 이러한 응용 모듈(380)을 포함하는 전자 장치(100)는 다양한 OS를 시스템의 OS로서 사용할 수 있다. 이러한 OS는 API(Application Program Interface: 361)에 하이 레벨 명령어를 제공하여 각 응용 모듈(380)의 동작을 제어한다. 본 발명의 전자 장치(100)는 API(361)로부터 제공되는 하이 레벨 명령어에 따라 대응하는 각 응용 모듈(380)을 식별하고, 하이 레벨 명령어를 디코딩하여 해당하는 곳으로 제공하는 하이 레벨 명령어 처리부(362)를 포함한다. 응용모듈 제어부(370)는 하이 레벨 명령어 처리부(362)로부터 제공된 명령어에 따라 응용 모듈(380)의 동작을 제어한다. 즉, 하이 레벨 명령어 처리부(362)는 API를 통하여 제공된 하이 레벨 명령어에 따라 여기에 대응하는 응용 모듈(380)이 존재하는지를 식별하고, 대응되는 응용 모듈(380)이 존재하는 경우에 해당하는 응용 모듈(380)에서 인식할 수 있는 명령어로 디코딩하여 해당하는 매핑부에 전송하거나 메시지 전송을 제어한다. 따라서, 응용 모듈 제어부(370)는 동작 인식 모듈(381)과, 음성 인식 모듈(382) 및 명령 처리 모듈(383)에 대한 매핑부(371, 373, 375)와 인터페이스부(372, 374, 376)를 각각 포함한다.

[0087] 동작 인식 모듈 매핑부(371)는 하이 레벨 명령어 처리부(362)로부터 전자 장치(100)에 부착된 카메라를 통해 사용자의 동작을 인식하고 인식 결과를 처리하기 위한 하이 레벨의 명령어를 제공받아, 동작 인식 모듈(381)에서

처리할 수 있는 디바이스 레벨로 매핑하고, 동작 인식 모듈 인터페이스부(372)를 통하여 동작 인식 모듈(381)로 제공한다. 음성 인식 모듈 매핑부(373) 및 음성 인식 모듈 인터페이스부(374)는 전자 장치(100)에 부착된 마이크를 통해 사용자의 음성 명령을 인식하고 이를 처리하는 부분이다. 즉, 음성 인식 모듈 매핑부(373)는 하이 레벨 명령어 처리부(362)로부터 음성 인식 모듈(382)을 이용하기 위한 하이 레벨 명령어를 제공받아 이를 디바이스 레벨의 명령어로 매핑시키고, 음성 인식 모듈 인터페이스부(374)를 통하여 음성 인식 모듈(382)로 제공한다. 명령 처리 모듈(383)은 음성 인식 모듈(382)을 통해 전달된 사용자의 음성 명령을 분석하고 이를 처리하는 부분이다. 명령 처리 모듈 매핑부(375)는 사용자의 음성 명령을 실행하기 위하여, 하이 레벨 명령어 처리부(362)를 통하여 인가된 하이 레벨 명령어를 제공받아 명령 처리 모듈(383)이 인식할 수 있는 디바이스 레벨의 명령어로 매핑시킨다. 디바이스 레벨의 명령어는 명령 처리 모듈 인터페이스부(376)를 통하여 명령 처리 모듈(383)로 제공된다. 이에 따라, 명령 처리 모듈(383)이 실행되어 사용자의 움직임에 의한 동작 인식 값과 메모리 시스템(150)에 저장된 트리거 기준과의 비교 결과가 제공되며, 동작 인식 값이 트리거 기준의 범위 내에 있는 경우에 전자 장치의 음성 인식 모드가 활성화되고, 음성 인식 모듈(382)에 따라 분석된 사용자의 음성 명령을 수행하게 된다.

[0088] 이러한 기능을 수행하도록 구성된 API(361)에 대한 세부적인 멤버 함수를 좀 더 상세하게 기술하면 다음과 같다. 개방 API(Open API)는 전자 장치(100)의 요청에 따라 장치의 온/오프, 잠금 화면의 표시, 동작 인식, 음성 인식 및 사용자의 음성 명령 처리에 관한 응용 모듈의 세션을 개방하기 위하여 사용된다. 폐쇄 API(Close API)는 사용된 응용 모듈에 대한 세션을 종료하는데 사용된다. 검색 API(Retrieve API)는 전자 장치(100)에서 호출 가능한 응용 모듈을 검색하는데 사용된다. 상태 API(Status API)는 각 응용 모듈의 동작 상태를 판단하기 위해 사용된다. 초기화 API(Initialize API)는 응용 모듈의 호출할 수 있도록 초기화하는 기능을 가진다. 리스트 API(List API)는 전자 장치(100)를 통해 사용자에게 제공되는 기능이나 동작, 동작 인식이나 음성 인식을 통하여 실행가능한 명령어 리스트 등을 확인하는데 사용된다. 기록 API(Register API)는 사용자가 전자 장치(100)를 통해 입력한 동작이나 음성 명령 등의 정보를 기록하는 기능을 가진다. 삭제 API(Unregister API)는 사용자가 입력한 동작이나 음성 명령어 등의 정보를 삭제하는데 사용된다. 결국, 사용되는 응용 모듈이나 메시지 전송 형태에 따라 이와 같은 개별적인 API가 실행되고, 그에 따라 사용자의 동작 인식이나 음성 인식, 명령 실행을 위한 응용 모듈의 이용이 가능할 것이다.

[0090] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명의 그 기술적 사상이나 필수적인 특징들이 변경되지 않고서 다른 구체적인 형태로 실시될 수 있다는 것으로 이해할 수 있을 것이다. 그러므로, 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

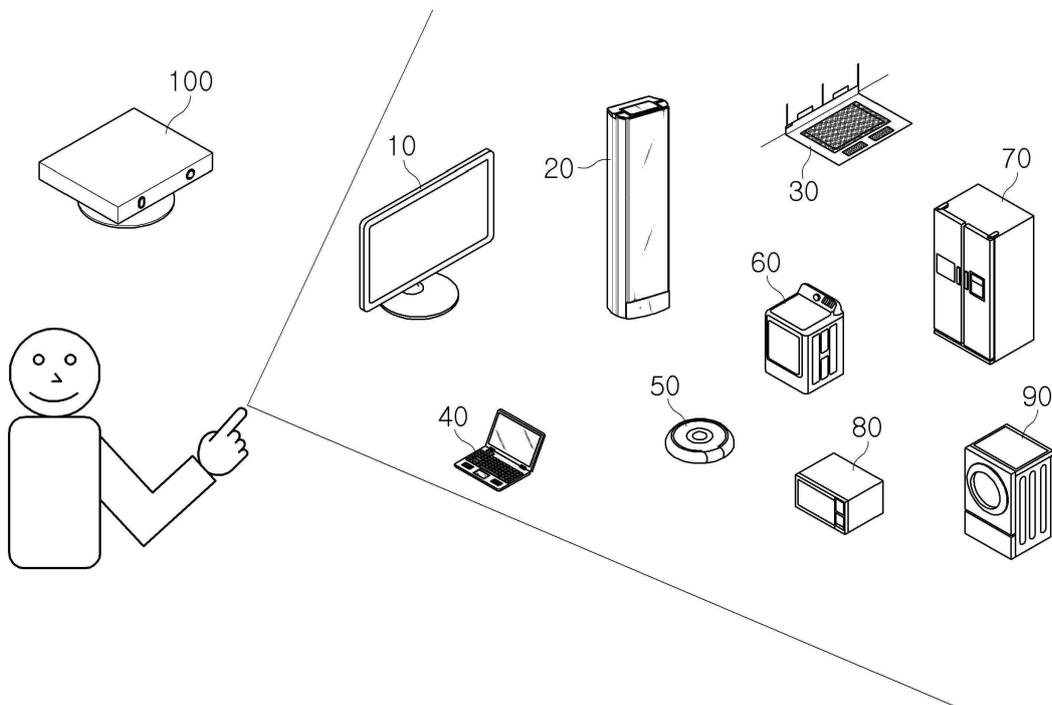
**부호의 설명**

- [0091]
- |                |                |
|----------------|----------------|
| 10: TV         | 20: 에어컨        |
| 30: 조리 기기      | 40: 노트북        |
| 50: 로봇 청소기     | 60: 식기 세척기     |
| 70: 냉장고        | 80: 전자레인지      |
| 90: 세탁기        | 100: 전자 장치     |
| 110: 입력 장치     | 112: 카메라       |
| 114: 마이크       | 120: 출력 장치     |
| 130: 센서부       | 132: 동작 센서     |
| 134: 음성 센서     | 140: 제어부       |
| 142: CPU       | 144: 영상 신호 처리부 |
| 146: 음성 신호 처리부 | 150: 메모리 시스템   |
| 152: 메인 메모리    | 154: 보조 메모리    |

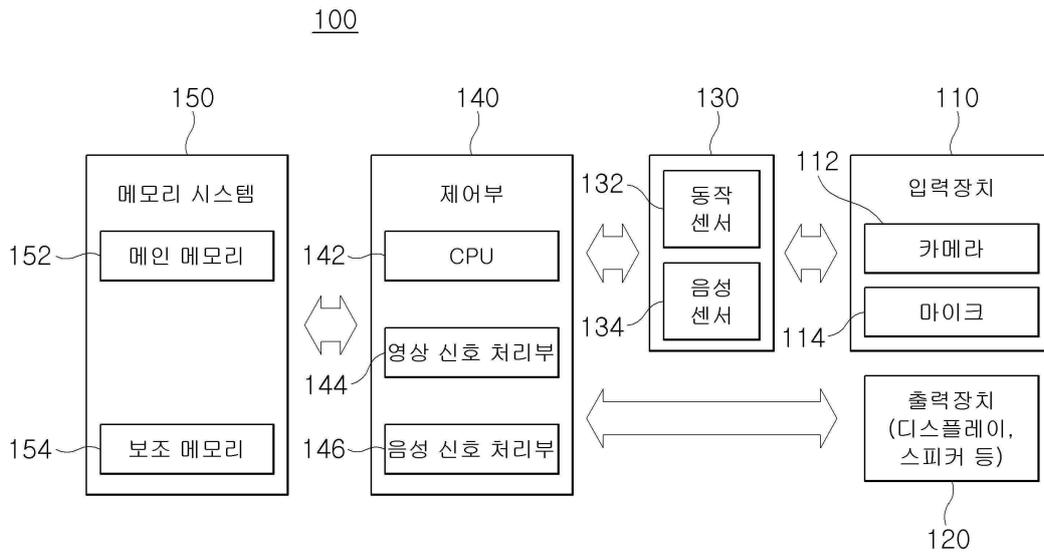
- 361: API
- 362: 하이레벨 명령어 처리부
- 370: 응용모듈 제어부    380: 응용모듈
- 371: 동작 인식 모듈 매핑부
- 372: 동작 인식 모듈 인터페이스부
- 373: 음성 인식 모듈 매핑부
- 374: 음성 인식 모듈 인터페이스부
- 375: 명령 처리 모듈 매핑부
- 376: 명령 처리 모듈 인터페이스부
- 381: 동작 인식 모듈
- 382: 음성 인식 모듈
- 383: 명령 처리 모듈

**도면**

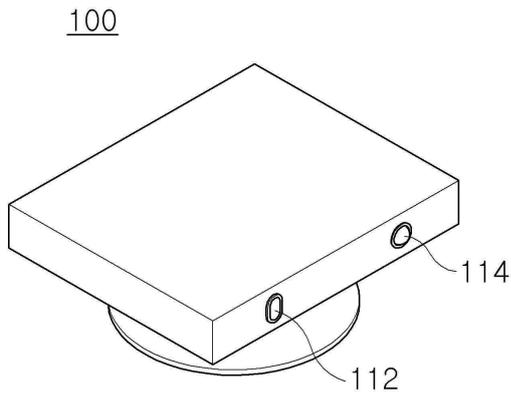
**도면1**



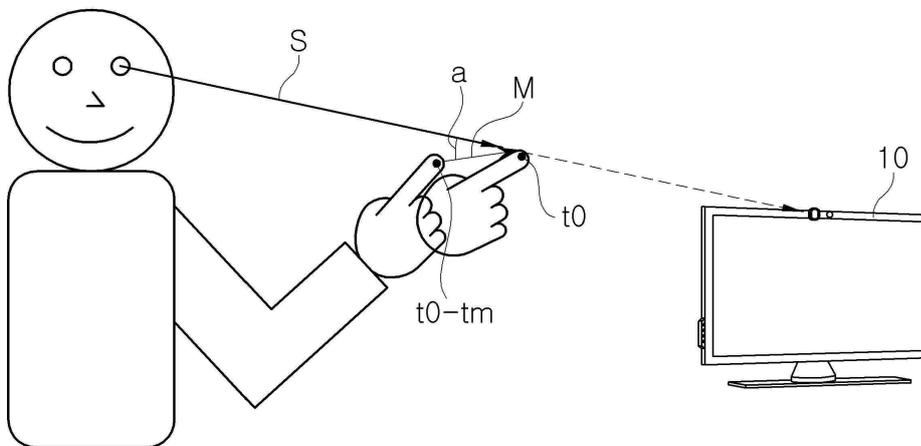
도면2



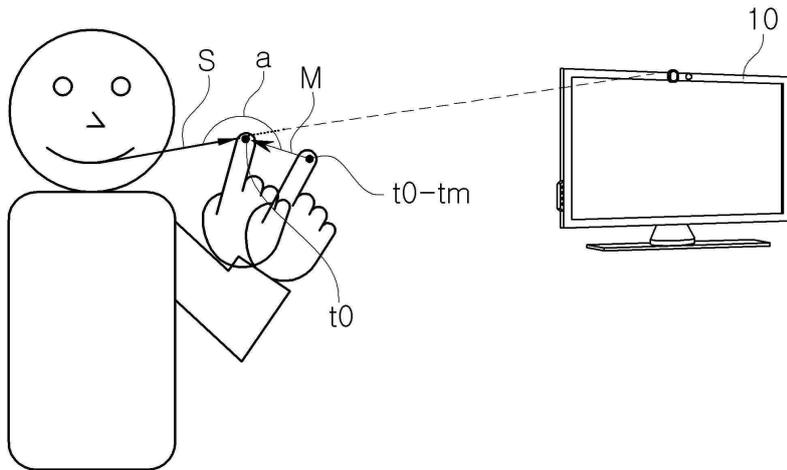
도면3



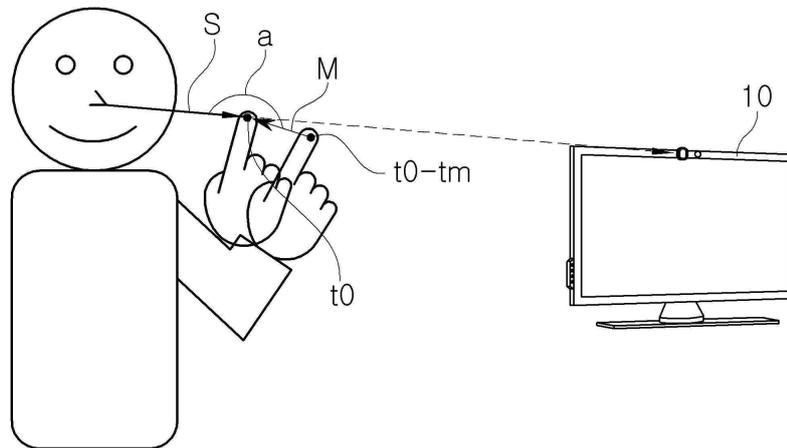
도면4



도면5



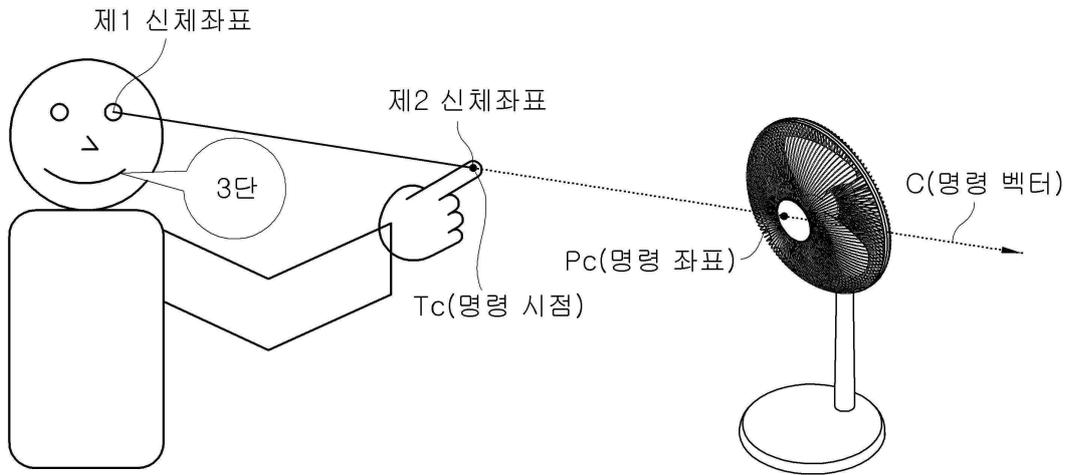
도면6



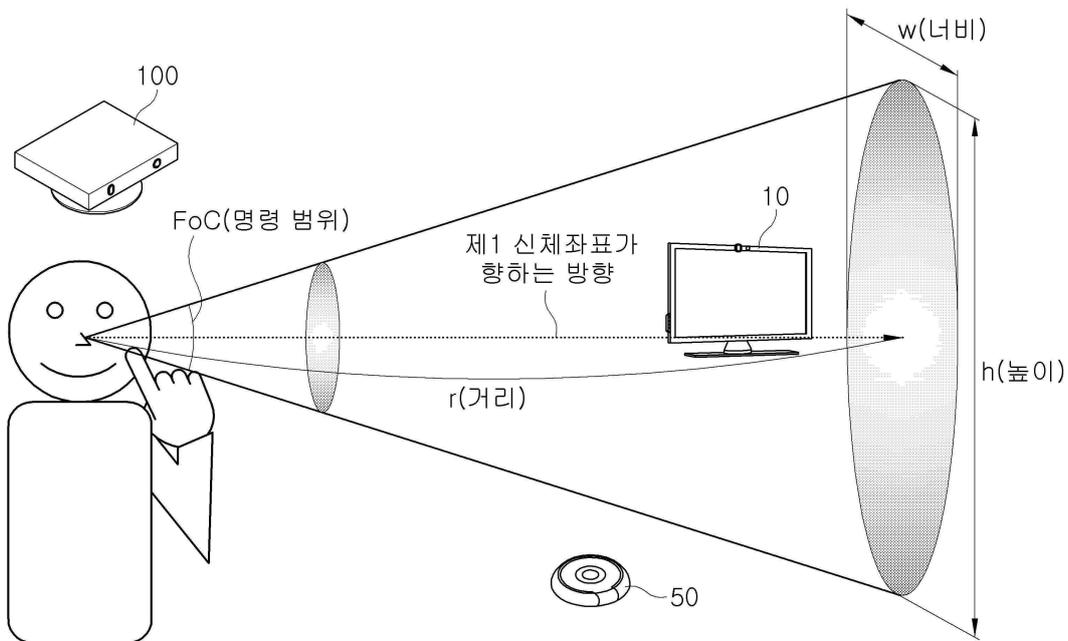
도면7

제1 신체 좌표	유효시간 ( $\Delta t_n$ )	임계조건	이동 벡터( $\vec{M}$ )-연결 벡터( $\vec{S}$ )의 방향 차( $a^\circ$ )	이동벡터의 크기( $ M $ )	연결벡터의 크기( $ S $ )	손포즈	명령벡터
한 눈	0.2sec	제1 임계조건	$-40^\circ < a^\circ < 40^\circ$	3cm 이상		1 finger (검지)	제1 명령벡터 Pointing
		제2 임계조건	-	1cm 미만	10~100cm	1 finger (검지)	
코 또는 입	0.2sec	제1 임계조건	$140^\circ < a^\circ < 220^\circ$	3cm 이상		1 finger (검지)	제2 명령벡터 Hush
		제2 임계조건	-	1cm 미만	5cm 미만	1 finger (검지)	

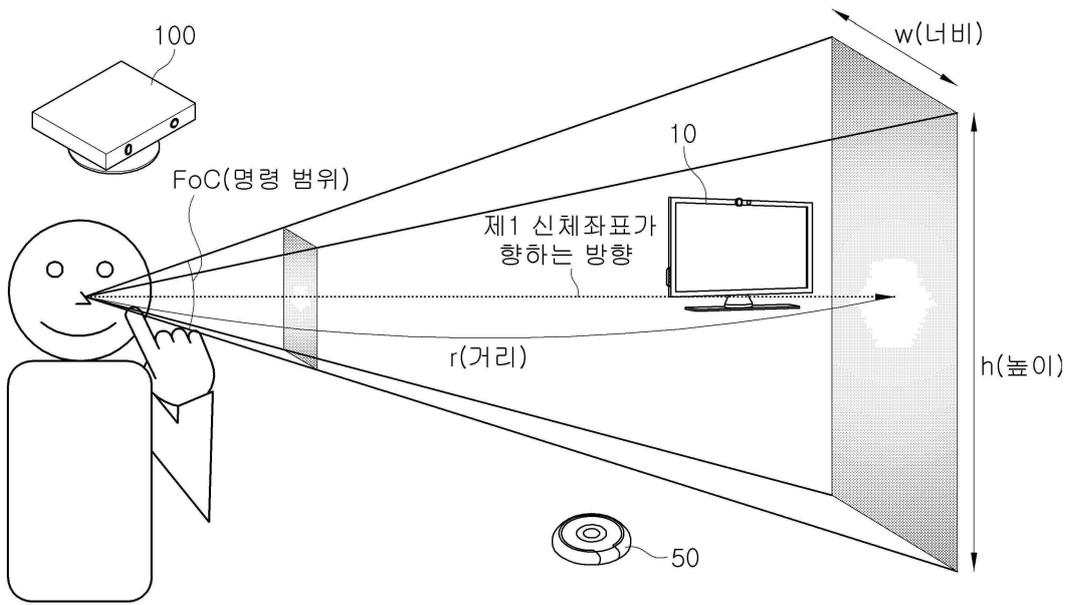
도면8



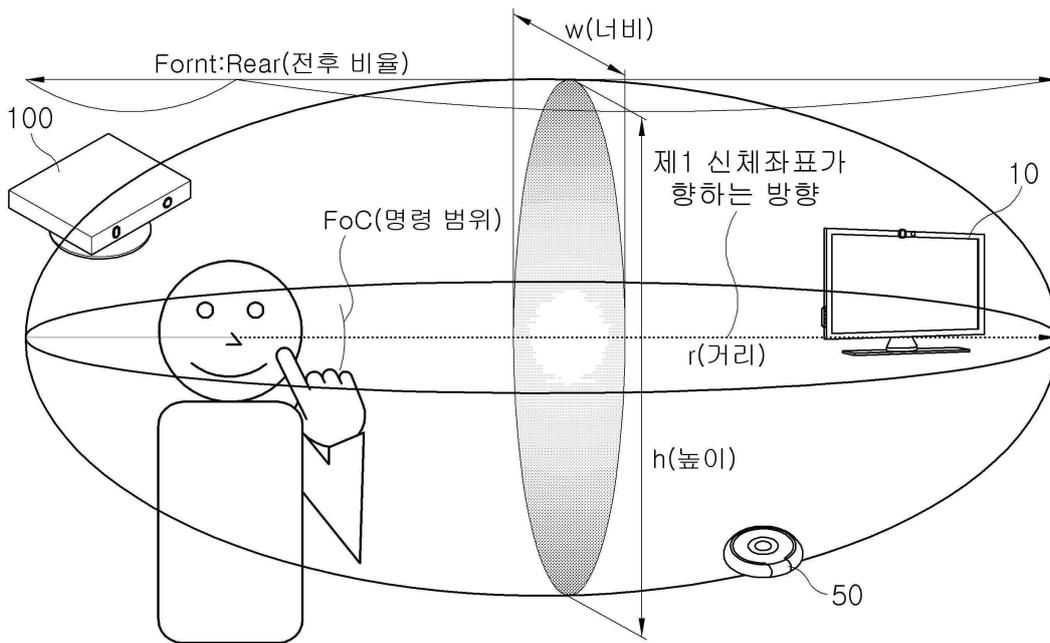
도면9



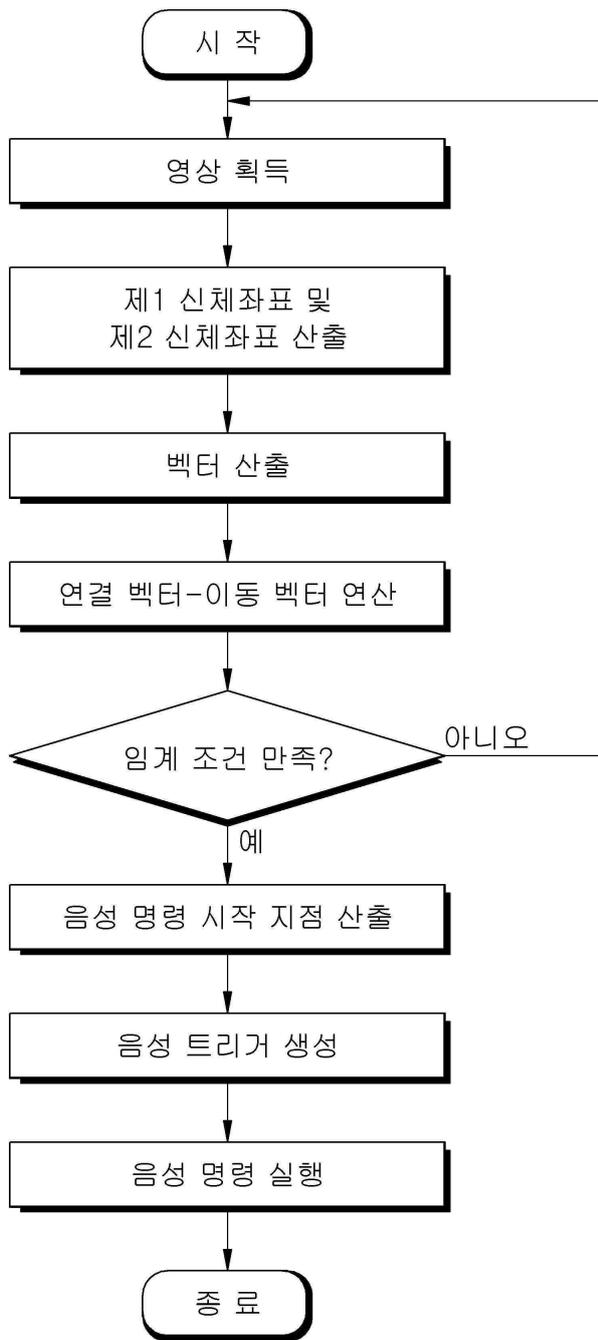
도면10



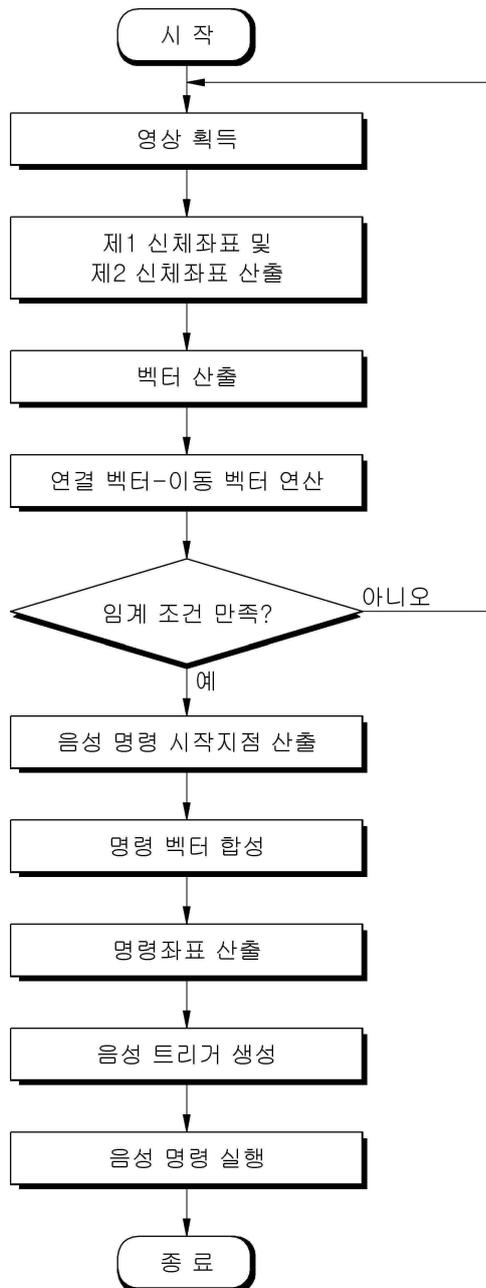
도면11



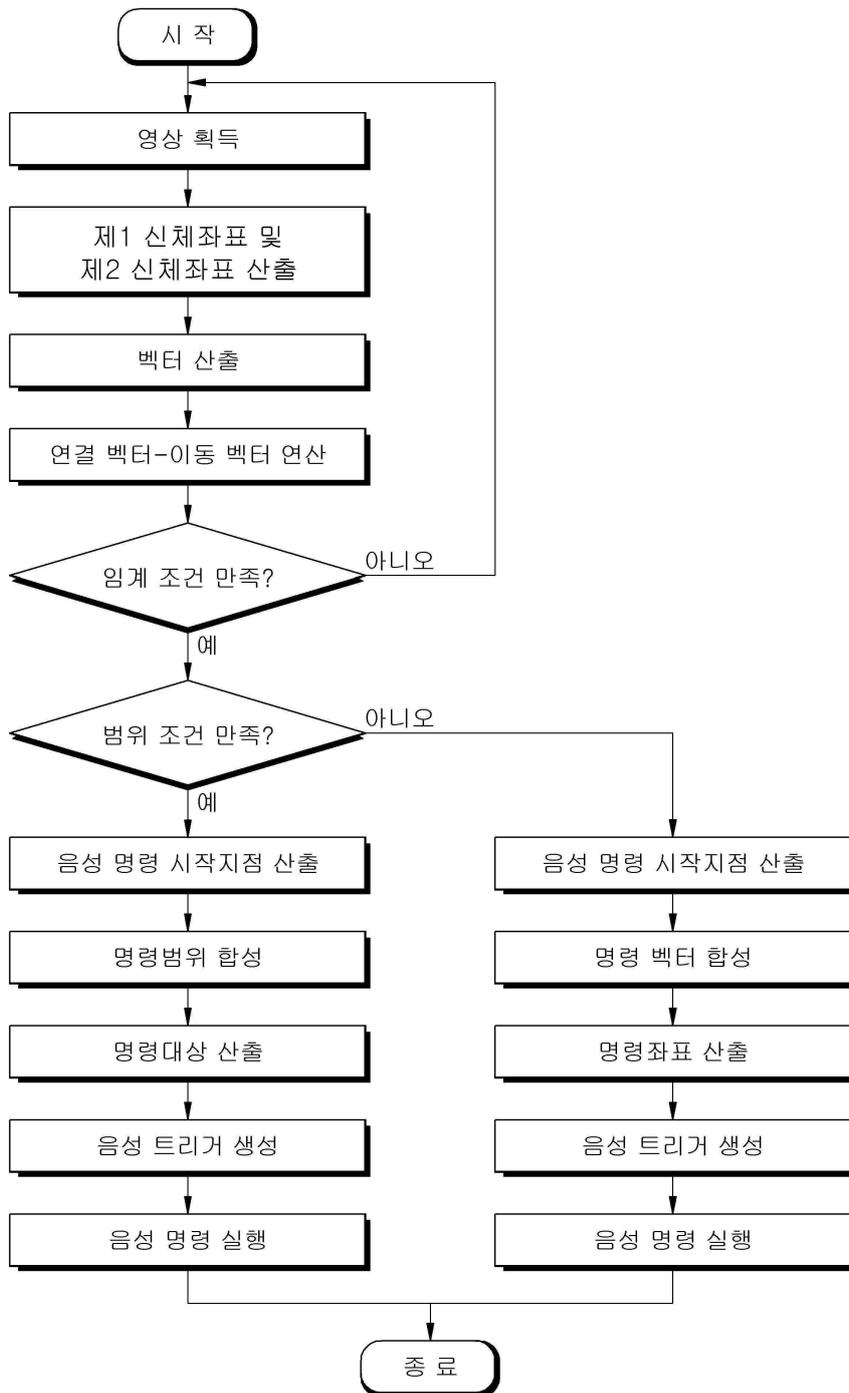
도면12



도면13



도면14



도면15

