



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0031459
(43) 공개일자 2020년03월24일

(51) 국제특허분류(Int. Cl.)
G10L 25/63 (2013.01) G06F 3/01 (2006.01)
G10L 15/26 (2006.01)
(52) CPC특허분류
G10L 25/63 (2013.01)
G06F 3/01 (2013.01)
(21) 출원번호 10-2018-0110500
(22) 출원일자 2018년09월14일
심사청구일자 2018년09월14일

(71) 출원인
엘지전자 주식회사
서울특별시 영등포구 여의대로 128 (여의도동)
(72) 발명자
신용경
서울특별시 서초구 양재대로11길 19 LG전자 특허
센터
문윤지
서울특별시 서초구 양재대로11길 19 LG전자 특허
센터
(뒷면에 계속)
(74) 대리인
박병창

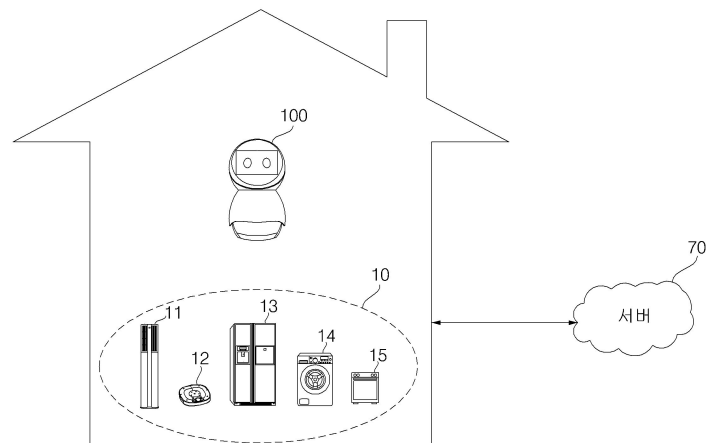
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 감성 인식기 및 이를 포함하는 로봇, 서버

(57) 요약

본 발명의 일 측면에 따른 감성 인식기는, 복수의 유니모달 입력 및, 복수의 유니모달 입력에 기초한 멀티모달 입력으로 감성 정보를 인식하도록 학습되어, 복수의 유니모달 입력 각각에 대한 감성 인식 결과와 멀티모달 입력에 대한 감성 인식 결과를 포함하는 복합 감성 인식 결과를 출력함으로써, 더욱 정확하게 사용자의 감성을 인식할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2013.01)

G10L 15/26 (2013.01)

(72) 발명자

박준우

서울특별시 서초구 양재대로11길 19 LG전자 특허센터

조택일

서울특별시 서초구 양재대로11길 19 LG전자 특허센터

명세서

청구범위

청구항 1

유니모달 입력 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 복수의 모달별 인식기를 포함하는 유니모달 전처리기; 및,

상기 복수의 모달별 인식기의 출력 데이터들을 병합하고, 병합된 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 멀티모달 인식기;를 포함하고,

상기 복수의 모달별 인식기 각각의 감성 인식 결과와 상기 멀티모달 인식기의 감성 인식 결과를 포함하는 복합 감성 인식 결과를 출력하는 것을 특징으로 하는 감성 인식기.

청구항 2

제1항에 있어서,

입력 데이터를 각 유니모달 별로 분리하여 상기 복수의 유니모달 입력 데이터를 생성하는 모달 분리기;를 더 포함하는 감성 인식기.

청구항 3

제1항에 있어서,

상기 복수의 유니모달 입력 데이터는,

상기 사용자를 포함하는 동영상 데이터에서 분리된 영상 유니모달 입력 데이터, 음성 유니모달 입력 데이터, 텍스트 유니모달 입력 데이터를 포함하는 것을 특징으로 하는 감성 인식기.

청구항 4

제3항에 있어서,

상기 텍스트 유니모달 입력 데이터는, 상기 동영상 데이터에서 분리된 음성을 텍스트로 변환한 데이터인 것을 특징으로 하는 감성 인식기.

청구항 5

제1항에 있어서,

상기 복수의 모달별 인식기는, 각각 입력되는 유니모달 입력 데이터의 입력 특성에 대응하는 인공신경망을 포함하는 것을 특징으로 하는 감성 인식기.

청구항 6

제1항에 있어서,

상기 멀티모달 인식기는, 상기 복수의 모달별 인식기가 출력하는 특징점 벡터들을 결합하는 병합기, 및, 상기 병합기의 출력 데이터에 포함된 상기 사용자의 감성 정보를 인식하도록 학습된 멀티모달 감성 인식기를 포함하는 것을 특징으로 하는 감성 인식기.

청구항 7

제1항에 있어서,

상기 복수의 모달별 인식기 각각의 감성 인식 결과는, 소정 개수의 기설정된 감성 클래스(class) 별 확률들을 포함하는 것을 특징으로 하는 감성 인식기.

청구항 8

제1항에 있어서,

상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 소정 기준에 따라 최종 감성 인식 결과를 출력하는 후처리기;를 더 포함하는 감성 인식기.

청구항 9

제8항에 있어서,

상기 후처리기는, 상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 모달별 인식기들의 감성 인식 결과들 중에서 상기 멀티모달 인식기의 감성 인식 결과와 일치하는 감성 인식 결과를 상기 최종 감성 인식 결과로 출력하는 것을 특징으로 하는 감성 인식기.

청구항 10

제8항에 있어서,

상기 후처리기는, 상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 복합 감성 인식 결과 중 2개의 감성 클래스를 포함하는 모순 감성을 상기 최종 감성 인식 결과로 출력하는 것을 특징으로 하는 감성 인식기.

청구항 11

제10항에 있어서,

상기 후처리기는, 상기 멀티모달 인식기의 감성 인식 결과 중 확률이 가장 높은 2개의 감성 클래스를 상기 모순 감성으로 선택하는 것을 특징으로 하는 감성 인식기.

청구항 12

제1항 내지 제11항 중 어느 한 항의 감성 인식기;를 포함하는 로봇.

청구항 13

사용자를 포함하는 동영상 데이터를 서버로 송신하고, 상기 서버로부터 상기 사용자의 감성 인식 결과를 복수개 포함하는 복합 감성 인식 결과를 수신하는 통신부; 및,

상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 일치하지 않는 2이상의 인식 결과를 조합하여 사용자의 감정을 확인하는 질문을 발화하는 음향 출력부;를 포함하는 로봇.

청구항 14

제13항에 있어서,

상기 수신된 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 소정 기준에 따라 최종 감성 인식 결과를 출력하는 후처리기;를 더 포함하는 로봇.

청구항 15

제14항에 있어서,

상기 후처리기는, 상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 복합 감성 인식 결과 중 2개의 감성 클래스를 포함하는 모순 감성을 상기 최종 감성 인식 결과로 출력하는 것을 특징으로 하는 로봇.

청구항 16

제15항에 있어서,

상기 후처리기는, 상기 복합 감성 인식 결과 중 확률이 가장 높은 2개의 감성 클래스를 상기 모순 감성으로 선

택하는 것을 특징으로 하는 로봇.

청구항 17

제13항에 있어서,

상기 서버는,

유니모달 입력 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 복수의 모달별 인식기를 포함하는 유니모달 전처리기, 및,

상기 복수의 모달별 인식기의 출력 데이터들을 병합하고, 병합된 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 멀티모달 인식기를 포함하고,

상기 복수의 모달별 인식기 각각의 감성 인식 결과와 상기 멀티모달 인식기의 감성 인식 결과에 기초하는 복합 감성 인식 결과를 상기 로봇으로 송신하는 것을 특징으로 하는 로봇.

청구항 18

제1항 내지 제11항 중 어느 한 항의 감성 인식기;를 포함하는 서버.

청구항 19

제18항에 있어서,

상기 서버는 로봇으로부터 화상 통화 데이터를 수신하고 상기 수신된 화상 통화 데이터에 포함된 사용자의 감성 인식 결과를 송신하는 것을 특징으로 하는 서버.

발명의 설명

기술 분야

[0001] 본 발명은 감성 인식기 및 이를 포함하는 로봇과 서버에 관한 것으로서, 더욱 상세하게는 사용자의 다양한 감성(emotion)을 인식할 수 있는 감성 인식기 및 이를 포함하는 로봇과 서버에 관한 것이다.

배경 기술

[0002] 로봇은 산업용으로 개발되어 공장 자동화의 일부분을 담당하여 왔다. 최근에는 로봇을 응용한 분야가 더욱 확대되어, 의료용 로봇, 우주 항공 로봇 등이 개발되고, 일반 가정에서 사용할 수 있는 가정용 로봇도 만들어지고 있다.

[0003] 로봇 이용의 증가에 따라, 단순한 기능 수행을 넘어서 사용자를 이해하고 교감하면서, 다양한 정보, 재미, 서비스를 제공할 수 있는 로봇에 대한 요구가 많아지고 있다.

[0004] 로봇 분야 뿐만 아니라 다양한 분야에서 사람의 감성을 인식하고 그에 따른 치료, 서비스를 제공하는 데 관심이 증가하고 있다. 이에 따라, 사람의 감성을 인식하는 방법에 관한 연구도 활발하게 이루어지고 있다.

[0005] 예를 들어, 사용자는 자신의 얼굴 등을 이용하여 고유의 캐릭터를 생성하고 이용하기도 한다. 선행 문헌 1(미국 특허등록공보 US 9262688B1)은 퍼지 집합을 이용하는 소정 알고리즘에 따라 멀티미디어 데이터로부터 감정 또는 표현을 인식하는 방법, 시스템에 관한 내용을 기술하고 있다.

[0006] 하지만, 이러한 선행 문헌에서 분석기 모듈은 후보 감정 또는 표현 데이터베이스로부터 최종적으로 1개의 감정 또는 표현을 선택하여 결과를 출력하게 된다.

[0007] 1개의 감정값만 출력하는 것은 감성과 관련된 정확하고 다양한 데이터를 획득하여 감성 기반의 서비스를 제공하는데 불충분할 수 있다. 또한, 입력 데이터 별로 감정의 상이함을 판단하는 것은 불가능하고, 여러 소스의 획득 데이터를 활용하더라도 최초 설정된 소스별 가중치에 영향을 많이 받는 한계가 있다.

발명의 내용

해결하려는 과제

- [0009] 본 발명의 목적은 사용자 감성(emotion)을 인식할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.
- [0010] 본 발명의 목적은 딥러닝으로 학습된 인공지능을 이용하여 사용자의 감성을 더욱 정확하게 인식할 수 있는 감성 인식 방법을 제공함에 있다.
- [0011] 본 발명의 목적은 사용자 감성(emotion)을 인식하고 감성기반의 서비스를 제공할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.
- [0012] 본 발명의 목적은 복수의 감성 인식 결과를 출력하고, 감성 인식 결과들을 조합하여 다양한 방식으로 이용할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.
- [0013] 본 발명의 목적은 복수의 감성 인식 결과에 기초하여 사용자와 대화함으로써 사용자와 교감할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.
- [0014] 본 발명의 목적은 유니모달과 멀티모달 감성 인식을 개별적으로 수행하고 복수의 인식 결과를 상호보완적으로 이용할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.
- [0015] 본 발명의 목적은 복합적인 감성을 인식할 수 있는 감성 인식기 및 이를 포함하는 로봇, 서버를 제공함에 있다.

과제의 해결 수단

- [0016] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 감성 인식기 및 이를 포함하는 로봇, 서버는, 사용자와 관련된 데이터를 획득하고, 획득된 사용자와 관련된 데이터에 기초하여 감성 정보를 인식하여, 감성기반의 서비스를 제공할 수 있다.
- [0017] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 감성 인식기는, 복수의 유니모달 입력 및, 복수의 유니모달 입력에 기초한 멀티모달 입력으로 감성 정보를 인식하도록 학습되어, 복수의 유니모달 입력 각각에 대한 감성 인식 결과와 멀티모달 입력에 대한 감성 인식 결과를 포함하는 복합 감성 인식 결과를 출력함으로써, 더욱 정확하게 사용자의 감성을 인식할 수 있다.
- [0018] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 감성 인식기는, 입력 데이터를 각 유니모달 별로 분리하여 복수의 유니모달 입력 데이터를 생성하는 모달 분리를 더 포함하여, 입력 데이터로부터 필요한 복수의 입력 데이터를 생성할 수 있다.
- [0019] 복수의 유니모달 입력 데이터는, 사용자를 포함하는 동영상 데이터에서 분리된 영상 유니모달 입력 데이터, 음성 유니모달 입력 데이터, 텍스트 유니모달 입력 데이터를 포함할 수 있고, 텍스트 유니모달 입력 데이터는, 상기 동영상 데이터에서 분리된 음성을 텍스트로 변환한 데이터일 수 있다.
- [0020] 복수의 모달별 인식기는, 각각 입력되는 유니모달 입력 데이터의 입력 특성에 대응하는 인공신경망을 포함함으로써, 개별 인식 결과의 정확성을 높일 수 있다. 또한, 멀티모달 인식기는, 순환신경망(recurrent neural networks)을 포함할 수 있다.
- [0021] 멀티모달 인식기는, 복수의 모달별 인식기가 출력하는 특징점 벡터들을 결합하는 병합기, 및, 병합기의 출력 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 멀티모달 감성 인식기를 포함할 수 있다.
- [0022] 복수의 모달별 인식기 각각의 감성 인식 결과와 멀티모달 인식기의 감성 인식 결과는, 소정 개수의 기설정된 감성 클래스(class) 별 확률들을 포함할 수 있다.
- [0023] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 감성 인식기 또는 로봇은, 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 소정 기준에 따라 최종 감성 인식 결과를 출력하는 후처리기를 더 포함할 수 있다.
- [0024] 이 경우에, 후처리기는, 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 모달별 인식기들의 감성 인식 결과들 중에서 멀티모달 인식기의 감성 인식 결과와 일치하는 감성 인식 결과를 최종 감성 인식 결과로 출력할 수 있다.

- [0025] 또는, 후처리는, 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 복합 감성 인식 결과 중 2개의 감성 클래스를 포함하는 모순 감성을 최종 감성 인식 결과로 출력할 수 있다. 이 경우에, 후처리는, 멀티모달 인식기의 감성 인식 결과 중 확률이 가장 높은 2개의 감성 클래스를 모순 감성으로 선택할 수 있다.
- [0026] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 로봇은 상술한 감성 인식기를 포함할 수 있다.
- [0027] 또한, 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 로봇은 화상 통화 상대방의 감성을 인식할 수 있다.
- [0028] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 로봇은 사용자를 포함하는 동영상 데이터를 서버로 송신하고, 서버로부터 사용자의 감성 인식 결과를 복수개 포함하는 복합 감성 인식 결과를 수신하는 통신부, 및, 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 일치하지 않는 2이상의 인식 결과를 조합하여 사용자의 감정을 확인하는 질문을 발화하는 음향 출력부를 포함할 수 있다.
- [0029] 한편, 감성 인식기가 복합 감성 인식 결과 중 2개의 감성 클래스를 포함하는 모순 감성을 최종 감성 인식 결과로 출력하는 경우에, 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 로봇은 2개의 감성 클래스를 조합하여 사용자의 감정을 확인하는 질문을 발화하여, 사용자와 대화할 수 있다.
- [0030] 상기 또는 다른 목적을 달성하기 위해 본 발명의 일 측면에 따른 서버는 상술한 감성 인식기를 포함할 수 있다. 서버는 로봇으로부터 화상 통화 데이터를 수신하고 상기 수신된 화상 통화 데이터에 포함된 사용자의 감성 인식 결과를 송신할 수 있다.

발명의 효과

- [0032] 본 발명의 실시예들 중 적어도 하나에 의하면, 사용자 감성(emotion)을 인식하고 감성기반의 서비스를 제공할 수 있다.
- [0033] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 딥러닝으로 학습된 인공지능을 이용하여 사용자의 감성을 더욱 정확하게 인식할 수 있다.
- [0034] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복수의 감성 인식 결과를 출력하고, 감성 인식 결과들을 조합하여 다양한 방식으로 이용할 수 있다.
- [0035] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복수의 감성 인식 결과에 기초하여 사용자와 대화함으로써 사용자와 교감하며 더욱 정확하게 사용자의 감성을 인식할 수 있다.
- [0036] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 유니모달과 멀티모달 감성 인식을 개별적으로 수행하고 복수의 인식 결과를 상호보완적으로 이용함으로써, 사용자의 감성을 더욱 정확하게 인식할 수 있다.
- [0037] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복합적인 감성을 인식할 수 있어, 사용자의 만족도와 편의성을 높일 수 있다.
- [0038] 한편, 그 외의 다양한 효과는 후술될 본 발명의 실시예에 따른 상세한 설명에서 직접적 또는 암시적으로 개시될 것이다.

도면의 간단한 설명

- [0040] 도 1은 본 발명의 일 실시예에 따른 로봇을 포함하는 로봇 시스템 구성도이다.
- 도 2는 본 발명의 일 실시예에 따른 로봇의 외관을 도시하는 정면도이다.
- 도 3은 본 발명의 일 실시예에 따른 로봇의 간략한 내부 블록도의 일예이다.
- 도 4는 본 발명의 일 실시예에 따른 서버의 간략한 내부 블록도의 일예이다.
- 도 5는 본 발명의 일 실시예에 따른 감성 인식기의 간략한 내부 블록도의 일예이다.

도 6은 본 발명의 실시예에 따른 감성 인식에 관한 설명에 참조되는 도면이다.

도 7 내지 도 9는 본 발명의 실시예에 따른 유니모달 감성 인식에 관한 설명에 참조되는 도면이다.

도 10은 본 발명의 실시예에 따른 멀티모달 감성 인식에 관한 설명에 참조되는 도면이다.

도 11은 본 발명의 일 실시예에 따른 감성 인식 결과를 예시한 도면이다.

도 12는 본 발명의 일 실시예에 따른 감성 인식 후처리에 관한 설명에 참조되는 도면이다.

도 13은 본 발명의 일 실시예에 따른 로봇의 감성 교류 사용자 경험에 관한 설명에 참조되는 도면이다.

도 14는 본 발명의 일 실시예에 따른 감성 인식기의 동작 방법을 도시한 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0041] 이하에서는 첨부한 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다. 그러나 본 발명이 이러한 실시예에 한정되는 것은 아니며 다양한 형태로 변형될 수 있음은 물론이다.
- [0042] 도면에서는 본 발명을 명확하고 간략하게 설명하기 위하여 설명과 관계없는 부분의 도면을 생략하였으며, 명세서 전체를 통하여 동일 또는 극히 유사한 부분에 대해서는 동일한 도면 참조부호를 사용한다.
- [0043] 한편, 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 단순히 본 명세서 작성의 용이함만이 고려되어 부여되는 것으로서, 그 자체로 특별히 중요한 의미 또는 역할을 부여하는 것은 아니다. 따라서, 상기 "모듈" 및 "부"는 서로 혼용되어 사용될 수도 있다.
- [0044] 또한, 본 명세서에서, 다양한 요소들을 설명하기 위해 제1, 제2 등의 용어가 이용될 수 있으나, 이러한 요소들은 이러한 용어들에 의해 제한되지 아니한다. 이러한 용어들은 한 요소를 다른 요소로부터 구별하기 위해서만 이용된다.
- [0045] 도 1은 본 발명의 일 실시예에 따른 로봇을 포함하는 로봇 시스템 구성도이다.
- [0046] 도 1을 참조하면, 본 발명의 일 실시예에 따른 로봇 시스템은, 하나 이상의 로봇(100), 및, 통신 모듈을 구비하여 다른 기기, 로봇(100), 서버(70) 등과 통신하거나 네트워크에 접속할 수 있는 홈 어플라이언스(home appliance, 10)들을 포함하여 구성될 수 있다.
- [0047] 예를 들어, 홈 어플라이언스(10)에는 통신 모듈을 구비한 공기조화기(11), 로봇 청소기(12), 냉장고(13), 세탁기(14), 조리기기(15) 등이 해당될 수 있다.
- [0048] 한편, 홈 어플라이언스(10)가 구비하는 통신 모듈은 와이파이(wi-fi) 통신 모듈일 수 있으며, 본 발명은 통신 방식에 한정되지 않는다.
- [0049] 또는, 홈 어플라이언스(10)는 다른 종류의 통신 모듈을 구비하거나 복수의 통신 모듈을 구비할 수 있다. 예를 들어, 홈 어플라이언스(10)는 NFC 모듈, 지그비(zigbee) 통신 모듈, 블루투스(Bluetooth™) 통신 모듈 등을 포함할 수 있다.
- [0050] 홈 어플라이언스(10)는 와이파이(wi-fi) 통신 모듈 등을 통해 소정 서버(70)와 연결 가능하고, 원격 모니터링, 원격 제어 등 스마트 기능을 지원할 수 있다.
- [0051] 본 발명의 일 실시예에 따른 로봇 시스템은, 스마트 폰(smart phone), 태블릿(Tablet) PC 등 휴대 단말기(미도시)를 포함할 수 있다.
- [0052] 사용자는 휴대 단말기를 통하여 로봇 시스템 내의 홈 어플라이언스(10, 20)에 관한 정보를 확인하거나 홈 어플라이언스(10)를 제어할 수 있다.
- [0053] 한편, 사용자가 가정 내에서 홈 어플라이언스(10)를 제어하거나 소정 정보를 확인하고자 하는 경우에도 휴대 단말기를 반드시 이용해야 하는 것은 불편할 수 있다.
- [0054] 예를 들어, 사용자가 휴대 단말기의 현재 위치를 모르거나 다른 장소에 있는 경우에 다른 방식으로 홈 어플라이언스(10)를 제어할 수 있는 수단이 있는 것이 더 효율적이다.
- [0055] 본 발명의 일 실시예에 따른 로봇(100)은 사용자의 음성 입력을 수신할 수 있고, 그에 따라 직접 홈 어플라이언스(10)를 제어하거나, 서버(70)를 경유하여 홈 어플라이언스(10)를 제어할 수 있다.

- [0056] 이에 따라, 사용자는 방, 거실 등에 배치된 로봇(100) 외에 다른 기기를 조작하지 않고서도 홈 어플라이언스(10)를 제어할 수 있다.
- [0057] 한편, 본 발명의 일 실시예에 따른 로봇 시스템은, 복수의 사물인터넷(IoT) 기기(미도시)를 포함하여 구성될 수 있다. 따라서, 로봇 시스템은, 홈 어플라이언스(10), 로봇(100), 사물인터넷(IoT) 기기들을 포함할 수 있다.
- [0058] 본 발명의 일 실시예에 따른 로봇 시스템은, 네트워크를 구성하는 통신 방식에 한정되지 않는다.
- [0059] 예를 들어, 홈 어플라이언스(10), 로봇(100), 사물인터넷(IoT) 기기들은, 유/무선 공유기(미도시)를 통하여, 통신 연결될 수 있다.
- [0060] 또한, 로봇 시스템 내의 기기들은 각각 개별적으로 통신 연결되는 메쉬 토폴로지(mesh topology)를 구성할 수 있다.
- [0061] 로봇 시스템 내의 홈 어플라이언스(10)는 유/무선 공유기(미도시)를 경유하여 서버(70)나 로봇(100)과 통신할 수 있다.
- [0062] 또한, 로봇 시스템 내의 홈 어플라이언스(10)는 이더넷(Ethernet)에 의해서 서버(70)나 로봇(100)과 통신할 수 있다.
- [0063] 본 발명의 일 실시예에 따른 로봇 시스템은, 게이트웨이(Gateway) 등 네트워크 장치를 더 포함할 수 있다. 또는, 가정(home) 내에 구비되는 로봇(100) 중 적어도 하나가 상술하는 게이트웨이(Gateway) 기능을 포함하도록 구성될 수 있다.
- [0064] 로봇 시스템에 포함되는 홈 어플라이언스(10)는, 기기들 간에 직접 또는 게이트웨이를 경유하여 네트워크 연결될 수 있다.
- [0065] 또한, 홈 어플라이언스(10)는, 직접 또는 게이트웨이를 경유하여 서버(70)와 통신 가능하도록 네트워크 연결될 수 있다.
- [0066] 한편, 상기 게이트웨이는 이더넷(Ethernet)에 의해서 서버(70)나 휴대 단말기(50)와 통신할 수 있다.
- [0067] 또한, 상기 게이트웨이는 유/무선 공유기를 경유하여 서버(70)나 로봇(100)과 통신할 수 있다.
- [0068] 홈 어플라이언스(10)는 서버(70) 및/또는 게이트웨이에 기기 작동 상태 정보, 설정값 정보 등을 전송할 수 있다.
- [0069] 사용자는 휴대 단말기, 또는 로봇(100)을 통하여 로봇 시스템 내의 홈 어플라이언스(10)에 관한 정보를 확인하거나 홈 어플라이언스(10)를 제어할 수 있다.
- [0070] 한편, 상기 서버(70) 및/또는 게이트웨이는 로봇(100) 등을 통한 사용자 명령 입력이 있거나, 로봇 시스템 내의 홈 어플라이언스(10)에서 발생하는 특정 이벤트(event)에 대응하여, 홈 어플라이언스(10)들을 제어하는 신호를 각 기기로 전송할 수 있다.
- [0071] 한편, 실시예에 따라서는 상기 게이트웨이는 디스플레이, 오디오 출력부 등 출력 수단을 구비할 수 있다.
- [0072] 이 경우에, 상기 디스플레이와 오디오 출력부는 상기 게이트웨이에 저장되어 있거나, 수신되는 신호에 기초한 영상과 오디오를 출력할 수 있다. 예를 들어, 상기 게이트웨이에 저장된 음악 파일을 재생하여 상기 오디오 출력부를 통하여 출력할 수 있다.
- [0073] 또한, 상기 디스플레이와 오디오 출력부는 상기 게이트웨이의 동작과 관련된 영상, 오디오 정보를 출력할 수 있다.
- [0074] 상기 서버(70)는 홈 어플라이언스(10), 로봇(100), 기타 기기로부터 전달되는 정보를 저장 및 관리할 수 있다.
- [0075] 상기 서버(70)는 홈 어플라이언스의 제조사 또는 제조사가 서비스를 위탁한 회사가 운영하는 서버일 수 있다.
- [0076] 홈 어플라이언스(10)와 관련된 정보는 상기 로봇(100)으로 전송될 수 있고, 상기 로봇(100)은 상기 홈 어플라이언스(10) 관련 정보를 표시할 수 있다.
- [0077] 상기 홈 어플라이언스(10)는 상기 로봇(100)으로부터 정보를 전달받거나 명령을 수신할 수 있다. 이때, 상기 홈 어플라이언스(10)는 상기 서버(70)로 각종 정보를 전송할 수 있고, 상기 서버(70)가 상기 홈 어플라이언스(10)로부터 수신한 정보의 일부 또는 전부를 상기 로봇(100)으로 전송할 수 있다.

- [0078] 그리고, 상기 서버(70)는 상기 홈 어플라이언스(10)로부터 수신한 정보 그 자체 또는 수신한 정보를 가공하여 상기 로봇(100)으로 전송할 수 있다.
- [0079] 한편, 도 1에서는 상기 서버(70)가 하나인 경우를 예시하였지만, 본 발명은 이에 한정되지 않고, 본 발명에 따른 시스템은 2개 이상의 서버와 연동하여 동작할 수 있다.
- [0080] 예를 들어, 상기 서버(70)는 음성 인식 및 처리를 위한 제1 서버와 홈 어플라이언스 제어 등 홈 어플라이언스 관련 서비스 제공을 위한 제2 서버를 포함할 수 있다.
- [0081] 실시예에 따라서는, 상기 제1 서버와 상기 제2 서버는, 복수의 서버로 정보, 기능이 분산되어 구성될 수도 있고, 하나의 통합 서버로 구성될 수도 있을 것이다.
- [0082] 예를 들어, 상기 음성 인식 및 처리를 위한 제1 서버는 음성 신호에 포함되는 단어들을 인식하는 음성 인식 서버와 음성 신호에 포함되는 단어들을 포함하는 문장의 의미를 인식하는 자연어 처리 서버로 구성될 수 있다.
- [0083] 또는, 상기 서버(70)는 감성 인식 및 처리를 위한 서버와 홈 어플라이언스 제어 등 홈 어플라이언스 관련 서비스 제공을 위한 서버를 포함할 수 있다. 이 경우에도, 감성 인식 및 처리를 위한 서버가 복수의 서버로 정보, 기능이 분산되어 구성될 수도 있고, 하나의 통합 서버로 구성될 수도 있다.
- [0084] 도 2는 본 발명의 일 실시예에 따른 로봇의 외관을 도시하는 정면도이고, 도 3은 본 발명의 일 실시예에 따른 로봇의 간략한 내부 블록도의 일예이다.
- [0085] 도 2와 도 3을 참조하면, 로봇(100)은, 외관을 형성하고 그 내부에 각종 부품을 수납하는 본체(101, 102)를 포함한다.
- [0086] 본체(101, 102)는 로봇(100)을 구성하는 각종 부품들이 수용되는 공간을 형성하는 바디(body, 101)와 상기 바디(101)의 하측에 배치되어 상기 바디(101)를 지지하는 지지부(102)를 포함할 수 있다.
- [0087] 또한, 로봇(100)은 본체(101, 102)의 상측에 배치되는 헤드(head, 110)를 포함할 수 있다. 헤드(110)의 전면에는 영상을 표시할 수 있는 디스플레이(182)가 배치될 수 있다.
- [0088] 본 명세서에서 전면 방향은 +y 축 방향을 의미하고, 상하 방향은 z축 방향, 좌우 방향은 x축 방향을 의미할 수 있다.
- [0089] 상기 헤드(110)는 x축을 중심으로 소정 각도 범위 내에서 회전할 수 있다.
- [0090] 이에 따라, 전면에서 봤을 때, 상기 헤드(110)는 사람이 고개를 상하 방향으로 끄덕거리는 것처럼 상하 방향으로 움직이는 노딩(Nodding) 동작이 가능하다. 예를 들어, 상기 헤드(110)는 사람이 머리를 상하 방향으로 끄덕거리는 것처럼 소정 범위 내에서 회전 후 원위치 복귀 동작을 1회 이상 수행할 수 있다.
- [0091] 한편, 실시예에 따라서는, 헤드(100) 중 사람의 안면에 대응할 수 있는 디스플레이(182)가 배치되는 전면 중 적어도 일부가 노딩되도록 구현될 수 있다.
- [0092] 따라서, 본 명세서에서는, 상기 헤드(110) 전체가 상하 방향으로 움직이는 실시예를 중심으로 기술하지만, 특별히 설명하지 않는 한, 헤드(110)가 상하 방향으로 노딩(Nodding)하는 동작은, 디스플레이(182)가 배치되는 전면 중 적어도 일부가 상하 방향으로 노딩하는 동작으로 대체 가능할 것이다.
- [0093] 상기 바디(101)는 좌우 방향으로 회전 가능하도록 구성될 수 있다. 즉, 상기 바디(101)는 z축을 중심으로 360도 회전 가능하도록 구성될 수 있다.
- [0094] 또한, 실시예에 따라서는, 상기 바디(101)도 x축을 중심으로 소정 각도 범위 내에서 회전가능하게 구성됨으로써, 상하 방향으로도 끄덕거리는 것처럼 움직일 수 있다. 이 경우에, 상기 바디(101)가 상하 방향으로 회전함에 따라, 상기 바디(101)가 회전하는 축을 중심으로 상기 헤드(110)도 함께 회전할 수 있다.
- [0095] 따라서, 본 명세서에서 헤드(110)가 상하 방향으로 노딩(Nodding)하는 동작은, 헤드(110) 자체가 소정 축을 중심으로 전면에서 봤을 때 상하 방향으로 회전하는 경우와 상기 바디(101)가 상하 방향으로 노딩함에 따라, 상기 바디(101)에 연결된 상기 헤드(110)가 함께 회전함으로써 노딩되는 경우를 모두 포함할 수 있다.
- [0096] 한편, 로봇(100)은 가정 내 콘센트에 연결되어 로봇(100) 내로 전원을 공급하는 전원 공급부(미도시)를 포함할 수 있다.
- [0097] 또는, 로봇(100)은 충전 가능한 배터리(미도시)를 구비하여 로봇(100) 내로 전원을 공급하는 전원 공급부(미도

시)를 포함할 수 있다. 실시예에 따라서는 전원 공급부(미도시)는 상기 배터리를 무선 충전하기 위한 무선전력 수신부를 구비할 수 있다.

- [0098] 한편, 로봇(100)은 본체(101, 102) 주변, 적어도 본체(101, 102) 전면을 중심으로 소정 범위를 촬영할 수 있는 영상 획득부(120)를 포함할 수 있다.
- [0099] 영상 획득부(120)는 본체(101, 102) 주변, 외부 환경 등을 촬영하는 것으로, 카메라 모듈을 포함할 수 있다. 상기 카메라 모듈은 디지털 카메라를 포함할 수 있다. 디지털 카메라는 적어도 하나의 광학렌즈와, 광학렌즈를 통과한 광에 의해 상이 맺히는 다수개의 광다이오드(photodiode, 예를 들어, pixel)를 포함하여 구성된 이미지센서(예를 들어, CMOS image sensor)와, 광다이오드들로부터 출력된 신호를 바탕으로 영상을 구성하는 디지털 신호 처리기(DSP: Digital Signal Processor)를 포함할 수 있다. 디지털 신호 처리기는 정지영상은 물론이고, 정지영상으로 구성된 프레임들로 이루어진 동영상을 생성하는 것도 가능하다.
- [0100] 이러한 카메라는 촬영 효율을 위해 각 부위별로 여러 개가 설치될 수도 있다. 바람직하게, 영상 획득부(120)는, 본체(101, 102) 전면의 영상을 획득하도록 헤드(110)의 전면에 구비되는 전면 카메라를 포함할 수 있다. 하지만, 영상 획득부(120)가 구비하는 카메라의 개수, 배치, 종류, 촬영범위가 반드시 이에 한정되어야 하는 것은 아니다.
- [0101] 영상 획득부(120)는, 로봇(100)의 전면 방향을 촬영할 수 있고, 사용자 인식용 영상을 촬영할 수 있다.
- [0102] 또한, 상기 영상 획득부(120)가 촬영하여 획득된 영상은 저장부(130)에 저장될 수 있다.
- [0103] 또한, 로봇(100)은 사용자의 음성 입력을 수신하는 음성 입력부(125)를 포함할 수 있다.
- [0104] 음성 입력부(125)는 아날로그 소리를 디지털 데이터로 변환하는 처리부를 포함하거나 처리부에 연결되어 사용자 입력 음성 신호를 서버(70) 또는 제어부(140)에서 인식할 수 있도록 데이터화할 수 있다.
- [0105] 음성 입력부(125)는 사용자 음성 입력 수신 of 정확도를 높이고, 사용자의 위치를 판별하기 위해, 복수의 마이크를 포함할 수 있다.
- [0106] 예를 들어, 음성 입력부(125)는 적어도 2이상의 마이크를 포함할 수 있다.
- [0107] 복수의 마이크(MIC)는, 서로 다른 위치에 이격되어 배치될 수 있고, 음성 신호를 포함한 외부의 오디오 신호를 획득하여 전기적인 신호로 처리할 수 있다.
- [0108] 한편, 입력 장치인 마이크는 음향을 발생시킨 음원, 사용자의 방향 추정을 위하여 최소 2개가 필요하며, 마이크 사이의 간격은 물리적으로 멀리 떨어져 있을수록 방향 검출의 해상도(각도)가 높다.
- [0109] 실시예에 따라서는 2개의 마이크가 상기 헤드(110)에 배치될 수 있다.
- [0110] 또한, 상기 헤드(110)의 후면에 2개의 마이크를 더 포함함으로써, 사용자의 3차원 공간상의 위치를 판별할 수 있다.
- [0111] 도 2를 참조하면, 로봇(100)은 전반적인 동작을 제어하는 제어부(140), 각종 데이터를 저장하는 저장부(130), 서버(70) 등 다른 기기와 데이터를 송수신하는 통신부(190)를 포함할 수 있다.
- [0112] 또한, 로봇(100)은 상기 헤드(110), 상기 바디(101)를 회전하는 구동부(160)를 더 포함할 수 있다. 구동부(160)는 바디(101) 및 헤드(110)를 회전 및/또는 이동시키는 복수의 구동 모터(미도시)를 포함할 수 있다.
- [0113] 제어부(140)는 로봇(100)을 구성하는 영상 획득부(120), 구동부(160), 디스플레이(182) 등을 제어하여, 로봇(100)의 동작 전반을 제어한다.
- [0114] 저장부(130)는 로봇(100)의 제어에 필요한 각종 정보들을 기록하는 것으로, 휘발성 또는 비휘발성 기록 매체를 포함할 수 있다. 기록 매체는 마이크로 프로세서(micro processor)에 의해 읽힐 수 있는 데이터를 저장한 것으로, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등을 포함할 수 있다.
- [0115] 또한, 제어부(140)는 통신부(190)를 통해 로봇(100)의 동작상태 또는 사용자 입력 등을 서버(70) 등으로 전송할 수 있다.
- [0116] 통신부(190)는 적어도 하나의 통신모듈을 포함하여 로봇(100)이 인터넷, 또는 소정의 네트워크에 연결되도록 한다.

- [0117] 또한, 통신부(190)는 홈 어플라이언스(10)에 구비되는 통신 모듈과 연결하여 로봇(100)과 홈 어플라이언스(10) 간의 데이터 송수신을 처리한다.
- [0118] 한편, 저장부(130)에는 음성 인식을 위한 데이터가 저장될 수 있고, 제어부(140)는 음성 입력부(125)를 통하여 수신되는 사용자의 음성 입력 신호를 처리하고 음성 인식 과정을 수행할 수 있다.
- [0119] 음성 인식 과정은 공지된 다양한 음성 인식 알고리즘을 사용할 수 있으므로, 본 명세서에서 음성 인식 과정에 대한 상세한 설명은 생략하기로 한다.
- [0120] 한편, 제어부(140)는 음성 인식 결과에 기초하여 로봇(100)이 소정 동작을 수행하도록 제어할 수 있다.
- [0121] 예를 들어, 제어부(140)는 음성 신호에 포함되는 명령이 소정 홈 어플라이언스의 동작을 제어하는 명령인 경우, 상기 음성 신호에 포함되는 명령에 기초한 제어 신호를 제어 대상 홈 어플라이언스로 전송하도록 제어할 수 있다.
- [0122] 또한, 제어부(140)는 음성 신호에 포함되는 명령이 소정 홈 어플라이언스의 동작을 제어하는 명령인 경우, 상기 로봇의 바디(101)를 제어 대상 홈 어플라이언스를 향하는 방향으로 회전하도록 제어할 수 있다.
- [0123] 한편, 음성 인식 과정은 로봇(100) 자체에서 실시되지 않고 서버(70)에서 수행될 수 있다.
- [0124] 이 경우에, 제어부(140)는 사용자 입력 음성 신호가 상기 서버(70)로 송신되도록 통신부(190)를 제어할 수 있다.
- [0125] 또는, 간단한 음성 인식은 로봇(100)이 수행하고, 자연어 처리 등 고차원의 음성 인식은 서버(70)에서 수행될 수 있다.
- [0126] 예를 들어, 기설정된 키워드를 포함하는 키워드 음성 입력이 수신되는 경우에, 로봇은 스탠바이(standby) 상태에서 동작 상태로 전환될 수 있다. 이 경우에, 로봇(100)은 키워드 음성 입력 여부까지의 음성 인식 과정만 수행하고, 이후의 사용자 음성 입력에 대한 음성 인식은 서버(70)를 통하여 수행할 수 있다.
- [0127] 한편, 실시예에 따라서는, 제어부(140)는 영상 획득부(120)를 통하여 획득되는 사용자 영상을 저장부(130)에 기 저장된 정보와 비교하여 등록된 사용자인지 여부를 판별할 수 있다.
- [0128] 또한, 제어부(140)는 등록된 사용자의 음성 입력에 대해서만 특정 동작을 수행하도록 제어할 수 있다.
- [0129] 한편, 상기 제어부(140)는, 상기 영상 획득부(120)를 통하여 획득된 사용자 영상 정보에 기초하여, 상기 바디(101) 및/또는 상기 헤드(111)를 회전하도록 제어할 수 있다.
- [0130] 이에 따라, 사용자와 로봇(100) 상호 간의 인터랙션(interaction)과 커뮤니케이션(communication)이 용이하게 할 수 있다.
- [0131] 한편, 로봇(100)은 출력부(180)를 포함하여, 소정 정보를 영상으로 표시하거나 음향으로 출력할 수 있다.
- [0132] 출력부(180)는 사용자의 명령 입력에 대응하는 정보, 사용자의 명령 입력에 대응하는 처리 결과, 동작모드, 동작상태, 에러상태 등을 영상으로 표시하는 디스플레이(182)를 포함할 수 있다.
- [0133] 상기 디스플레이(182)는 상술한 것과 같이 헤드(110)의 전면에 배치될 수 있다.
- [0134] 실시예에 따라서는, 상기 디스플레이(182)는 터치패드와 상호 레이어 구조를 이루어 터치스크린으로 구성될 수 있다. 이 경우에, 상기 디스플레이(182)는 출력 장치 이외에 사용자의 터치에 의한 정보의 입력이 가능한 입력 장치로도 사용될 수 있다.
- [0135] 또한, 출력부(180)는 오디오 신호를 출력하는 음향 출력부(181)를 더 포함할 수 있다. 음향 출력부(181)는 제어부(140)의 제어에 따라 경고음, 동작모드, 동작상태, 에러상태 등의 알림 메시지, 사용자의 명령 입력에 대응하는 정보, 사용자의 명령 입력에 대응하는 처리 결과 등을 음향으로 출력할 수 있다. 음향 출력부(181)는, 제어부(140)로부터의 전기 신호를 오디오 신호로 변환하여 출력할 수 있다. 이를 위해, 스피커 등을 구비할 수 있다.
- [0136] 도 2를 참조하면, 음향 출력부(181)는 헤드(110)의 좌우측면에 배치되어, 소정 정보를 음향으로 출력할 수 있다.
- [0137] 한편, 도 2에 예시된 로봇의 외관 및 구조는 예시적인 것으로 본 발명은 이에 한정되지 않는다. 예를 들어, 음

성 입력부(125), 영상 획득부(120), 음향 출력부(181)의 위치, 개수 등은 설계 사양에 따라 달라질 수 있다. 또한, 각 구성 부품의 회전 방향과 각도도 달라질 수 있다. 예를 들어, 도 2에서 예시된 로봇(100)의 회전 방향과 달리 로봇(100) 전체가 특정 방향으로 기울어지거나 흔들리는 동작도 가능하다.

- [0138] 한편, 본 발명의 일 실시예에 따른 로봇(100)은, 유선 또는 무선 인터넷 기능의 지원으로 인터넷 및 컴퓨터에 접속될 수 있다.
- [0139] 또한, 본 발명의 일 실시예에 따른 로봇(100)은 음성 및 화상 통화 기능도 수행 가능하며, 이러한 통화 기능은 VoIP(Voice over Internet Protocol)에 따른 인터넷망, 이동통신망 등을 이용하여 수행될 수 있다.
- [0140] 한편, 제어부(140)는, 화상 통화 시 화상 통화 상대방의 영상, 사용자의 영상을 사용자의 설정에 따라 표시하도록 디스플레이(182)를 제어할 수 있으며, 수신되는 화상 통화 상대방의 음성 신호에 기초한 음성을 음향 출력부(181)가 출력하도록 제어할 수 있다.
- [0141] 본 발명의 일 실시예에 따른 로봇 시스템은 화상 통화를 수행하는 2이상의 로봇을 포함할 수 있다.
- [0142] 도 4는 본 발명의 일 실시예에 따른 서버의 간략한 내부 블록도의 일예이다.
- [0143] 도 4를 참조하면, 서버(70)는, 통신부(72), 저장부(73), 인식기(74), 및 프로세서(71)를 구비할 수 있다.
- [0144] 프로세서(71)는, 서버(70)의 전반적인 동작을 제어할 수 있다.
- [0145] 한편, 서버(70)는, 로봇(100) 등 홈 어플라이언스 제조사가 운영하는 서버 또는 서비스 제공자가 운영하는 서버일 수 있고, 일종의 클라우드(Cloud) 서버일 수 있다.
- [0146] 통신부(72)는, 휴대 단말기, 로봇(100) 등 홈 어플라이언스, 게이트웨이 등으로부터 상태 정보, 동작 정보, 조작 정보 등 각종 데이터를 수신할 수 있다.
- [0147] 그리고 통신부(72)는 수신되는 각종 정보에 대응하는 데이터를 휴대 단말기, 로봇(100) 등 홈 어플라이언스, 게이트웨이 등으로 송신할 수 있다.
- [0148] 이를 위해, 통신부(72)는 인터넷 모듈, 이동 통신 모듈 등 하나 이상의 통신 모듈을 구비할 수 있다.
- [0149] 저장부(73)는, 수신되는 정보를 저장하고, 이에 대응하는 결과 정보 생성을 위한 데이터를 구비할 수 있다.
- [0150] 또한, 저장부(73)는, 머신 러닝에 사용되는 데이터, 결과 데이터 등을 저장할 수 있다.
- [0151] 인식기(74)는 상기 로봇(100) 등 홈 어플라이언스의 학습기 역할을 수행할 수 있다.
- [0152] 상기 인식기(74)는 인공신경망, 예를 들어, CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), DBN(Deep Belief Network) 등 심층신경망(Deep Neural Network: DNN)을 포함할 수 있고, 심층신경망을 학습할 수 있다.
- [0153] 한편, 상기 프로세서(71)는 설정에 따라 학습 후 상기 로봇(100) 등 홈 어플라이언스의 인공신경망 구조를 학습된 인공신경망 구조로 업데이트시키도록 제어할 수 있다.
- [0154] 또한, 상기 인식기(74)는, 인식용 입력 데이터를 입력받고, 입력 데이터에 포함된 사물, 공간, 감성의 속성을 인식하여, 그 결과를 출력할 수 있다. 이 경우에, 통신부(72)는 로봇(100)으로 인식 결과를 송신할 수 있다.
- [0155] 또한, 상기 인식기(74)는, 로봇(100)의 사용(Usage) 관련 데이터(Data)를 분석하고 학습하여 사용 패턴, 사용 환경 등을 인식하여, 그 결과를 출력할 수 있다. 이 경우에, 통신부(72)는 로봇(100)으로 인식 결과를 송신할 수 있다.
- [0156] 이에 따라, 로봇(100) 등 홈 어플라이언스 제품들은 서버(70)로부터 인식 결과를 수신하고, 수신된 인식 결과를 활용하여 동작할 수 있다.
- [0157] 상기 서버(70)는 사용자가 발화한 음성 입력 신호를 수신하여 음성 인식을 수행할 수 있다. 이를 위해, 상기 서버(70)는 음성 인식기를 포함할 수 있고, 음성 인식기 입력 데이터에 대하여 음성 인식을 수행하여 음성 인식 결과를 출력하도록 학습된 인공신경망을 포함할 수 있다.
- [0158] 한편, 상기 서버(70)는 음성 인식을 위한 음성 인식 서버를 포함할 수 있다. 또한, 음성 인식 서버도 음성 인식 과정 중 소정 과정을 분담하여 수행하는 복수의 서버를 포함할 수 있다. 예를 들어, 음성 인식 서버는, 음성 데이터를 수신하고, 수신한 음성 데이터를 텍스트(text) 데이터로 변환하는 자동 음성 인식(Automatic Speech

Recognition: ASR) 서버, 및, 상기 자동 음성 인식 서버로부터 상기 텍스트 데이터를 수신하고, 수신한 텍스트 데이터를 분석하여 음성 명령을 판별하는 자연어 처리(Natural Language Processing: NLP) 서버를 포함할 수 있다. 경우에 따라서, 음성 인식 서버는, 자연어 처리 서버가 출력한 텍스트 음성 인식 결과를 음성 데이터로 변환하여 다른 서버 또는 홈 어플라이언스로 송신하는 텍스트 음성 변환(Text to Speech: TTS) 서버를 더 포함할 수 있다.

- [0159] 상기 서버(70)는 입력 데이터에 대하여 감성 인식을 수행할 수 있다. 이를 위해, 상기 서버(70)는 감성 인식기를 포함할 수 있고, 감성 인식기는 입력 데이터에 대하여 감성 인식을 수행하여 감성 인식 결과를 출력하도록 학습된 인공신경망을 포함할 수 있다.
- [0160] 한편, 상기 서버(70)는 감성 인식을 위한 감성 인식 서버를 포함할 수 있다. 즉, 상기 서버(70) 중 적어도 하나는 감성 인식을 수행하는 감성 인식기를 구비하는 감성 인식 서버일 수 있다.
- [0161] 도 5는 본 발명의 일 실시예에 따른 감성 인식기의 간략한 내부 블록도의 일예이다.
- [0162] 도 5를 참조하면, 로봇(100) 또는 서버(70)가 구비하는 감성 인식기(74a)는, 감성 데이터를 입력 데이터(590)로 사용하여 딥러닝(Deep Learning)을 수행할 수 있다.
- [0163] 본 발명의 일 실시예에 따른 감성 인식기(74a)는 유니모달 입력 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 복수의 모달별 인식기(521, 522, 523)를 포함하는 유니모달 전처리기(520), 및, 상기 복수의 모달별 인식기(521, 522, 523)의 출력 데이터들을 병합하고, 병합된 데이터에 포함된 사용자의 감성 정보를 인식하도록 학습된 멀티모달 인식기(510)를 포함할 수 있다.
- [0164] 감성 데이터는 사용자의 감성에 대한 정보를 가지고 있는 감성 정보 데이터로, 감성 인식에 사용될 수 있는 감성 정보인 영상, 음성, 생체 신호 데이터를 포함할 수 있다. 바람직하게는, 상기 입력 데이터(590)는 사용자의 얼굴을 포함하는 비디오 데이터(Video)일 수 있고, 더욱 바람직하게는 상기 학습 데이터(590)는 사용자의 음성을 포함하는 오디오(Audio) 데이터를 더 포함할 수 있다.
- [0165] 감성은, 자극에 대하여 느낌이 일어나는 능력, 감각적 자극이나 인상을 받아들이는 마음의 성질로, 감성 공학에서는 환경의 변화나 외부로부터의 물리적인 자극에 대한 인간 내부의 고차원적인 심리적인 체험으로 쾌적감, 불쾌감 등 복합적인 감정으로 정의하고 있다.
- [0166] 본 명세서에서, 감성은 자극에 대하여 발생하는 쾌적감, 불쾌감 등의 감정을 의미할 수 있고, 감성은 N개의 대표적인 감정 상태 중 어느 하나로 인식될 수 있다. 이러한 N개의 대표적인 감정 상태를 감성 클래스로 명명할 수 있다.
- [0167] 예를 들어, 감성 인식기(74a)는 놀람(surprise), 행복(happiness), 슬픔(sadness), 싫음(displeasure), 분노(anger), 두려움(fear) 등 6개의 대표적인 감정 클래스를 인식하고 감성 인식 결과로 대표적인 감정 클래스 중 하나를 출력하거나, 6개의 대표적인 감정 클래스 별 확률값으로 출력할 수 있다.
- [0168] 또는, 본 발명의 일 실시예에 따른 감성 인식기(74a)가 인식하고 출력하는 감성으로 놀람, 행복, 슬픔, 싫음, 분노, 두려움과 같은 감성 클래스 외에 6개의 감정이 발생하지 않은 디폴트 감정 상태를 나타내는 중립(neutrality) 감성 클래스를 더 포함할 수 있다.
- [0169] 이 경우에, 감성 인식기(74a)는 놀람, 행복, 슬픔, 싫음, 분노, 두려움, 중립 중 선택된 어느 하나의 감성 클래스를 감성 인식 결과로 출력하거나, 놀람 x%, 행복 x%, 슬픔 x%, 싫음 x%, 분노 x%, 두려움 x%, 중립 x%와 같이 각 감성 클래스 별 확률값을 감성 인식 결과로 출력할 수 있다.
- [0170] 인지하고자 하는 감성을 딥러닝 학습시킨 인공지능 모델로 사용자의 감성을 인식하면, 결과값으로는 딥러닝 학습시 사용한 데이터의 태깅값으로 출력하게 된다.
- [0171] 한편, 실환경에서는 사용자의 감성이 최종적으로 1개의 감성으로 나올 수 없는 경우가 많다. 예를 들어, 사용자가 말로는 기쁨 감성을 표시하지만 얼굴 표정에서는 싫음 감성이 표현될 수 있다. 이와 같이, 사람들이 음성, 영상, 텍스트 등의 각 모달 별로 다른 감성을 출력하는 경우가 많다.
- [0172] 따라서, 사용자의 감성을 인식하여 최종 하나의 감성값으로 출력하거나 음성, 영상, 텍스트 각각의 다른 감성, 모순된 감성, 유사한 감성 등을 무시할 경우, 실제 사용자가 느끼는 감정과 다른 감정을 인식할 수 있다..
- [0173] 사용자의 외부로 드러나는 모든 정보를 기반으로 각각의 감성을 인지하고 관리하기 위하여, 본 발명의 일 실시

예에 따른 감성 인식기(74a)는, 음성, 영상, 텍스트 각각의 유니모달 별로 감성 인식이 가능하고, 멀티모달로도 감성 인식이 가능한 구조를 가질 수 있다.

- [0174] 즉, 본 발명의 일 실시예에 따른 감성 인식기(74a)는, 특정 시점에 입력되는 사용자의 감성을 유니모달 별로 인식하고, 동시에 멀티모달로 복합 인식할 수 있다.
- [0175] 한편, 복수의 모달별 인식기(521, 522, 523)는 각각 입력되는 한가지의 유니모달 입력 데이터를 인식하여 처리하는 것으로 유니모달 인식기로도 명명될 수 있다.
- [0176] 본 발명의 일 실시예에 따른 감성 인식기(74a)는 입력 데이터(590)를 각 유니모달 별로 분리하여 상기 복수의 유니모달 입력 데이터를 생성할 수 있다. 모달 분리기(530)는, 입력 데이터(590)를 복수의 유니모달 입력 데이터들로 분리할 수 있다.
- [0177] 여기서, 상기 복수의 유니모달 입력 데이터는, 상기 사용자를 포함하는 동영상 데이터에서 분리된 영상 유니모달 입력 데이터, 음성 유니모달 입력 데이터, 텍스트 유니모달 입력 데이터를 포함할 수 있다.
- [0178] 예를 들어, 입력 데이터(590)는 사용자가 촬영된 동영상 데이터일 수 있고, 동영상 데이터는 사용자의 얼굴 등이 촬영된 비디오 데이터와 사용자가 발화한 음성을 포함하는 오디오 데이터를 포함할 수 있다.
- [0179] 이 경우에, 모달 분리기(530)는 입력 데이터(590)에 포함되는 오디오 데이터의 내용을 텍스트 데이터로 변환한 텍스트 유니모달 입력 데이터(531)와 음성 톤, 크기, 높이 등 오디오 데이터의 음성(sound) 유니모달 입력 데이터로 분리할 수 있다.
- [0180] 상기 텍스트 유니모달 입력 데이터는, 상기 동영상 데이터에서 분리된 음성을 텍스트로 변환한 데이터일 수 있다. 경우에 따라서, 음성(sound) 유니모달 입력 데이터는 오디오 데이터의 음원 파일 그 자체이거나 음원 파일에서 노이즈 제거 등 전처리가 완료된 파일일 수 있다.
- [0181] 또한, 모달 분리기(530)는 입력 데이터(590)에 포함되는 비디오 데이터로부터 하나 이상의 얼굴 이미지 데이터를 포함하는 영상 유니모달 입력 데이터(533)를 분리할 수 있다.
- [0182] 한편, 분리된 유니모달 입력 데이터들(531, 532, 533)은, 각 유니모달 입력 데이터들(531, 532, 533)에 기초하여 사용자의 감성 정보를 인식하도록 학습된 복수의 모달별 인식기(521, 522, 523)를 포함하는 유니모달 전처리 기(520)로 입력될 수 있다.
- [0183] 예를 들어, 텍스트 유니모달 입력 데이터(531)는 텍스트를 학습 데이터로 사용하여 딥러닝을 수행하는 텍스트 감성 인식기(521)로 입력될 수 있다.
- [0184] 음성(sound) 유니모달 입력 데이터(532)는 음성 학습 데이터로 사용하여 딥러닝을 수행하는 음성(speech) 감성 인식기(522)로 입력될 수 있다.
- [0185] 하나 이상의 얼굴 이미지 데이터를 포함하는 영상 유니모달 입력 데이터(533)는 영상 학습 데이터로 사용하여 딥러닝을 수행하는 얼굴(face) 감성 인식기(523)로 입력될 수 있다.
- [0186] 텍스트 감성 인식기(521)는 텍스트로 변환된 STT(Sound to text) 데이터에 포함되는 어휘, 문장 구조 등을 인식하여 사용자의 감성을 인식할 수 있다. 예를 들어, 행복에 관련된 단어를 많이 사용하거나 행복의 정도를 강하게 표현하는 단어가 인식될수록 행복 감성 클래스에 대한 확률값이 다른 감성 클래스에 대한 확률값보다 높게 인식할 수 있다. 또는 텍스트 감성 인식기(521)는 인식된 텍스트에 대응하는 감성 클래스인 행복을 감성 인식 결과로 바로 출력할 수 있다.
- [0187] 또한, 텍스트 감성 인식기(521)는 감성 인식 결과와 함께 텍스트 특징점 벡터도 출력할 수 있다.
- [0188] 음성(speech) 감성 인식기(522)는 입력되는 음성 데이터의 특징점들을 추출한다. 이때, 상기 음성 특징점들은 음성의 톤(tone), 음량, 파형 등을 포함할 수 있다. 음성(speech) 감성 인식기(522)는 음성의 톤(tone) 등을 검출하여 사용자의 감성을 판별할 수 있다.
- [0189] 또한, 음성(speech) 감성 인식기(522)도 감성 인식 결과와 검출한 음성 특징점 벡터들을 출력할 수 있다.
- [0190] 얼굴 감성 인식기(523)는 입력된 이미지 데이터에서 사용자의 얼굴 영역을 검출하고, 얼굴 표정을 구성하는 특징점들인 표정 랜드마크(Landmark) 포인트 정보를 인식함으로써 사용자의 표정을 인식할 수 있다. 또한, 얼굴 감성 인식기(523)는 인식된 표정에 대응하는 감성 클래스 또는 감성 클래스별 확률값을 출력할 수 있고, 얼굴

특징점(표정 랜드마크 포인트) 벡터도 출력할 수 있다.

- [0191] 도 6은 본 발명의 실시예에 따른 감성 인식에 관한 설명에 참조되는 도면으로, 표정의 구성 요소들을 예시한 것이다.
- [0192] 도 6을 참조하면, 표정 랜드마크(Landmark) 포인트는 눈썹(61), 눈(62), 뺨(63), 이마(64), 코(65), 입(66), 턱(67) 등이 해당될 수 있다.
- [0193] 도 6에서의 표정 랜드마크(Landmark) 포인트들(61 내지 67)은 예시적인 것으로 종류와 숫자는 달라질 수 있다.
- [0194] 예를 들어, 눈썹(61), 눈(62), 입(66)과 같이 특징이 강한 소수의 표정 랜드마크 포인트만 이용하거나 사용자 별로 특정 표정을 지을 때 변화 정도가 큰 표정 랜드마크 포인트를 이용할 수도 있다.
- [0195] 한편, 얼굴 감성 인식기(523)는 표정 랜드마크(Landmark) 포인트들(61 내지 67)의 위치 및 모양에 기초하여 얼굴 표정을 인식할 수 있다.
- [0196] 얼굴 감성 인식기(523)는 표정 랜드마크 포인트들(61 내지 67) 중 적어도 일부가 포함되는 이미지 데이터로 딥러닝된 인공신경망을 포함하여 사용자의 얼굴 표정을 인식할 수 있다.
- [0197] 예를 들어, 사용자가 눈(62)을 크게 뜨고, 입(66)을 크게 벌리면 얼굴 감성 인식기(523)는 사용자의 감성을 감성 클래스들 중 행복으로 판별하거나 행복에 대한 확률값이 가장 높은 감성 인식 결과를 출력할 수 있다.
- [0198] 한편, 상기 복수의 모달별 인식기는, 각각 입력되는 유니모달 입력 데이터의 입력 특성에 대응하는 인공신경망을 포함할 수 있다. 또한, 멀티모달 감성 인식기(511)도 입력 데이터의 특성에 대응하는 인공신경망을 포함할 수 있다.
- [0199] 예를 들어, 영상 기반의 학습 및 인식을 수행하는 얼굴 감성 인식기(523)는 CNN(Convolutional Neural Network), 다른 감성 인식기들(521, 522)은 심층신경망(DNN), 멀티모달 감성 인식기(511)는 순환 신경망(Recurrent Neural Network: RNN)의 인공신경망을 포함할 수 있다.
- [0200] 모달별 감성 인식기(521, 522, 523)는 각각 입력되는 유니모달 입력 데이터(531, 532, 533)에 포함된 감성 정보를 인식하여 감성 인식 결과를 출력할 수 있다. 예를 들어, 모달별 감성 인식기(521, 522, 523)는 소정 개수의 기설정된 감성 클래스(class) 중 확률이 가장 큰 감성 클래스를 감성 인식 결과로 출력하거나 감성 클래스 별 확률값을 감성 인식 결과로 출력할 수 있다.
- [0201] 한편, 모달별 감성 인식기(521, 522, 523)는 각각의 딥러닝 구조에서 텍스트, 음성, 영상을 학습, 인식하고, 각 유니모달 별 특징점 벡터로 구성되는 중간 벡터값을 도출할 수 있다.
- [0202] 또한, 멀티모달 인식기(510)는, 각 음성, 영상, 텍스트의 중간 벡터값으로 멀티모달 딥러닝을 수행할 수 있다.
- [0203] 이와 같이, 모달별 감성 인식기(521, 522, 523)의 출력에 기초하여 멀티모달 인식기(510)의 입력이 생성되므로, 모달별 감성 인식기(521, 522, 523)들은 일종의 전처리기로 동작할 수 있다.
- [0204] 한편, 본 발명의 일 실시예에 따른 감성 인식기(74a)는, 3개의 모달별 감성 인식기(521, 522, 523)의 딥러닝 모델과 1개의 멀티모달 인식기(510)의 딥러닝 모델 총 4개의 딥러닝 모델을 사용할 수 있다.
- [0205] 한편, 멀티모달 인식기(510)는 복수의 모달별 인식기(521, 522, 523)가 출력하는 특징점 벡터들을 결합하는 병합기(512), 및, 상기 병합기(512)의 출력 데이터에 포함된 상기 사용자의 감성 정보를 인식하도록 학습된 멀티모달 감성 인식기(511)를 포함할 수 있다.
- [0206] 여기서, 병합기(512)는 복수의 모달별 인식기(521, 522, 523)의 출력 데이터들의 싱크를 맞추고, 특징점 벡터들을 결합(Vector Concatenation)하여 상기 멀티모달 감성 인식기(511)로 출력할 수 있다.
- [0207] 상기 멀티모달 감성 인식기(511)는 입력 데이터로부터 사용자의 감성 정보를 인식하여 감성 인식 결과를 출력할 수 있다.
- [0208] 예를 들어, 상기 멀티모달 감성 인식기(511)는 소정 개수의 기설정된 감성 클래스(class) 중 확률이 가장 큰 감성 클래스를 감성 인식 결과로 출력하거나 감성 클래스 별 확률값을 감성 인식 결과로 출력할 수 있다.
- [0209] 이에 따라, 본 발명의 일 실시예에 따른 감성 인식기(74a)는, 복수의 유니모달 감성 인식 결과와 하나의 멀티모달 감성 인식 결과를 출력할 수 있다.

- [0210] 본 발명의 일 실시예에 따른 감성 인식기(74a)는 복수의 유니모달 감성 인식 결과와 하나의 멀티모달 감성 인식 결과를 감성 클래스별 레벨(확률)로 출력할 수 있다.
- [0211] 예를 들어, 감성 인식기(74a)는 놀람, 행복, 중립, 슬픔, 싫음, 분노, 두려움의 감성 클래스 별 확률값을 출력할 수 있고, 확률값이 높을수록 인지된 감성 클래스일 가능성이 높다는 것을 의미할 수 있다. 여기서, 7종의 감성 클래스들의 확률값 총합은 100%가 될 것이다.
- [0212] 감성 인식기(74a)는, 상기 복수의 모달별 인식기 각각의 감성 인식 결과(521, 522, 523)와 상기 멀티모달 인식기(511)의 감성 인식 결과를 포함하는 복합 감성 인식 결과를 출력할 수 있다.
- [0213] 이에 따라, 로봇(100)은 유니모달 3개, 멀티모달 1개의 감성 인식 결과를 기반으로 감성교류 사용자 경험(UX)을 제공할 수 있다.
- [0214] 설정에 따라서, 감성 인식기(74a)는 복합 감성 인식 결과에서 다수를 차지하는 인식 결과, 확률값이 가장 높은 인식 결과를 최종 인식 결과로 출력할 수 있다. 또는, 복수의 감성 인식 결과를 수신한 로봇(100)의 제어부(140)가 소정 기준에 따라 최종 인식 결과를 판별할 수 있다.
- [0215] 본 발명에 따른 감성 인식기(74a)는 음성(음성톤 등), 영상(얼굴 표정 등), 텍스트(말의 내용 등) 각각의 감성을 레벨로 인식하여 관리할 수 있다. 이에 따라, 모달별로 감성교류 사용자 경험(UX)을 다르게 처리할 수 있다.
- [0216] 또한, 유니모달(음성, 영상, 텍스트)별 감성 인식 결과와 멀티모달 감성 인식 결과값을 한 시점 기준으로 동시에 출력할 수 있다. 한 시점에서 입력되는 음성, 영상, 텍스트를 복합적으로 감성을 인식할 수 있어, 멀티모달 감성에서 각 유니모달별로 모순된 감성을 인식하여 사용자의 감성 성향을 파악할 수 있다. 이에 따라, 일부 모달에서 부정적인 입력이 들어와도 전체적인 감성을 인식하여 사용자의 진짜 감정 상태인 긍정적인 입력에 대응하는 감성교류 사용자 경험(UX)을 제공할 수도 있다.
- [0217] 본 발명에 따르면 로봇(100)이 감성 인식기(74a)를 구비하거나 감성 인식기(74a)를 구비하는 서버(70)와 통신하여 사용자만의 유니모달별 감성을 파악할 수 있다.
- [0218] 또한, 본 발명에 따르면 사용자만의 감성 패턴 분석을 할 수 있고, 모달별 감성인식으로 감성케어(치료)에 활용 가능하다.
- [0219] 종래의 감성 방법은 입력 데이터의 모달별 인식 결과가 다른 모순된 감성의 경우 감성을 하나로 매핑하여 감성을 분석하기엔 어려움이 많았다.
- [0220] 하지만, 본 발명에 따르면, 여러 개의 입력과 출력을 통해 다양한 실생활 상황에 대해 가장 부합하게 처리할 수 있다.
- [0221] 낮은 성능을 보이는 입력 인식기를 보완하기 위해, 본 발명은 여러 개의 입력과 출력을 결합(Fusion) 방식으로 복수의 인식기(511, 521, 522, 523)가 서로 상호 보완하는 인식기 구조를 구성할 수 있다.
- [0222] 본 발명의 일 실시예에 따른 감성 인식기(74a)는, 음성을 소리와 뜻으로 분리하여, 영상, 음성 입력으로부터, 영상, 음성(사운드) 및 STT 총 3개의 입력을 만들 수 있다.
- [0223] 또한, 3개의 입력에 대해 각각 최적의 성능을 내기 위해 감성 인식기(74a)는 CNN(Convolutional Neural Network)와 Long Short-Term Memory(LSTM) 등 입력별로 다른 인공신경망 모델을 가지게 구성할 수 있다. 예를 들어, 영상 기반의 인식기(523)는 CNN 구조를 가지고, 멀티모달 감성 인식기(511)는 Long Short-Term Memory(LSTM) 구조를 가질 수 있다. 이에 따라, 각 입력 특성에 맞춤형 신경망을 구성할 수 있다.
- [0224] 각 입력에 대한 유니모달 인식기(521, 522, 523)의 출력은 7종 감성 클래스에 대한 확률값과 이 감성을 잘 표현하는 특징점들의 벡터값일 수 있다..
- [0225] 멀티모달 인식기(510)는, 단순히 3개의 입력에 대한 감성값을 통계적 방법으로 계산하는 것이 아닌 감성을 잘 표현하는 특징점들의 벡터값을 전결합 레이어와 LSTM을 통해 결합하여, 한 인식기가 어려워 하는 문제를 다른 인식기가 도와주는 형태로 성능 향상에 도움을 주며, 실생활에서 다양한 케이스들을 커버할 수 있다.
- [0226] 예를 들어, 얼굴을 인식하기 어려운 곳에서 말만 들리는 경우에도 본 발명의 일 실시예에 따른 감성 인식기(74a)는 음성 기반의 인식기(521, 522) 및 멀티모달 감성 인식기(511)가 사용자의 감성을 인식할 수 있다.
- [0227] 또한, 감성 인식기(74a)는 영상, 음성, 문자 데이터에 대한 각각의 인식 결과와 멀티모달 인식 결과를 융합함으

로써 사용자의 복잡한 감정 상태를 인식할 수 있기 때문에, 실제 생활에서의 다양한 상황에 대해서 감정 인식이 가능하다.

- [0228] 한편, 유니모달 전처리기(520)는 각각 입력되는 한가지의 유니모달 입력 데이터를 인식하여 처리하는 유니모달 인식기들(521, 522, 523)을 포함할 수 있다.
- [0229] 도 5를 참조하면, 유니모달 전처리기(520)는 텍스트 감정 인식기(521), 음성(speech) 감정 인식기(522), 얼굴 감정 인식기(523)를 포함할 수 있다.
- [0230] 이러한 유니모달 인식기들(521, 522, 523)은 사전에 학습되어 확보될 수 있다.
- [0231] 도 7 내지 도 9는 본 발명의 실시예에 따른 유니모달 감정 인식에 관한 설명에 참조되는 도면이다.
- [0232] 도 7은 얼굴 감정 인식기(523)의 유니모달 학습과정을 예시한 것이다.
- [0233] 도 7의 (a)를 참조하면, 영상 기반의 입력 데이터(710)에 기초하여, 얼굴 감정 인식을 위한 인공신경망(720)이 딥러닝을 수행할 수 있다.
- [0234] 영상 기반의 입력 데이터(710)는 비디오 데이터일 수 있고, 인공신경망(720)은 비디오 데이터로 학습하거나 비디오 데이터로부터 추출된 복수의 이미지 데이터로 학습을 수행할 수 있다.
- [0235] 도 7의 (a)에서는 5개의 대표 이미지(715)를 추출하여 학습하는 예를 도시하였으나, 본 발명은 이에 한정되지 않는다.
- [0236] 얼굴 감정 인식기(523)를 위한 인공신경망(720)은, 영상 기반의 학습 및 인식에 많이 사용되는 CNN(Convolutional Neural Network) 등일 수 있다.
- [0237] 이와 같이, 영상 처리에 장점이 있는 CNN 인공신경망(720)은 사용자의 얼굴을 포함하는 입력 데이터를 입력받아 감성을 인식하도록 학습될 수 있다.
- [0238] 도 7의 (b)를 참조하면, 얼굴 감정 인식기(523)를 위한 인공신경망(720)은 얼굴 이미지(725)가 입력되면, 입력된 얼굴 이미지(725)의 표정 랜드마크 포인트 등 특징점을 추출하고 사용자의 얼굴에 나타나는 감성을 인식할 수 있다.
- [0239] 얼굴 감정 인식기(523)를 위한 인공신경망(720)이 출력하는 감정 인식 결과(730)는 놀람, 행복, 슬픔, 싫음, 분노, 두려움, 중립 중 선택된 어느 하나의 감정 클래스일 수 있다. 더욱 바람직하게는 감정 인식 결과(730)는 놀람 x%, 행복 x%, 슬픔 x%, 싫음 x%, 분노 x%, 두려움 x%, 중립 x%와 같이 각 감정 클래스 별 확률값을 포함할 수 있다.
- [0240] 도 5를 참조하여 설명한 것과 같이, 모달별 감정 인식기(521, 522, 523)의 출력에 기초하여 멀티모달 인식기(510)의 입력이 생성되므로, 모달별 감정 인식기(521, 522, 523)들은 일종의 전처리기로 동작할 수 있다.
- [0241] 도 7의 (c)를 참조하면, 얼굴 감정 인식기(523)는 감정 인식 결과(730)를 출력할 뿐만 아니라, 입력된 얼굴 이미지에 기초하여 추출된 특징점 벡터인 히든 스테이트(hidden state, 740)를 출력할 수 있다.
- [0242] 도 8은 텍스트 감정 인식기(521)의 유니모달 학습과정을 예시한 것이다.
- [0243] 도 8의 (a)를 참조하면, 텍스트 기반의 입력 데이터(810)에 기초하여, 텍스트 감정 인식을 위한 인공신경망(820)이 딥러닝을 수행할 수 있다.
- [0244] 텍스트 기반의 입력 데이터(810)는 사용자가 발화한 음성(speech)을 텍스트(text)로 변환한 STT 데이터일 수 있고, 인공신경망(820)은 STT 데이터 또는 기타 텍스트 데이터로 학습을 수행할 수 있다.
- [0245] 텍스트 감정 인식기(521)를 위한 인공신경망(820)은, 딥러닝을 수행하는 심층신경망 DNN 중 하나일 수 있다.
- [0246] 도 8의 (b)를 참조하면, 텍스트 감정 인식기(521)를 위한 인공신경망(820)은 텍스트 데이터(825)가 입력되면, 입력된 텍스트 데이터(825)의 특징점을 추출하고 텍스트에 나타나는 감성을 인식할 수 있다.
- [0247] 텍스트 감정 인식기(521)를 위한 인공신경망(820)이 출력하는 감정 인식 결과(830)는 놀람, 행복, 슬픔, 싫음, 분노, 두려움, 중립 중 선택된 어느 하나의 감정 클래스일 수 있다. 더욱 바람직하게는 감정 인식 결과(830)는 놀람 x%, 행복 x%, 슬픔 x%, 싫음 x%, 분노 x%, 두려움 x%, 중립 x%와 같이 각 감정 클래스 별 확률값을 포함할 수 있다.

- [0248] 텍스트 감성 인식기(521)도 멀티모달 인식기(510)의 전처리기로 동작할 수 있다. 도 8의 (c)를 참조하면, 텍스트 감성 인식기(521)는 감성 인식 결과(830)를 출력할 뿐만 아니라, 입력된 텍스트 데이터에 기초하여 추출된 특징점 벡터인 히든 스테이트(840)를 출력할 수 있다.
- [0249] 도 9는 음성(speech) 감성 인식기(522)의 유니모달 학습과정을 예시한 것이다.
- [0250] 도 9의 (a)를 참조하면, 음성 기반의 입력 데이터(910)에 기초하여, 감성 인식을 위한 인공신경망(820)이 딥러닝을 수행할 수 있다.
- [0251] 음성 기반의 입력 데이터(910)는 사용자가 발화한 음성(speech)의 소리(sound)를 포함하는 데이터일 수 있고, 음원 파일 그 자체이거나 음원 파일에서 노이즈 제거 등 전처리가 완료된 파일일 수 있다.
- [0252] 인공신경망(920)은 음성 기반의 입력 데이터(910)에서 감성을 인식하도록 학습을 수행할 수 있다.
- [0253] 음성 감성 인식기(522)를 위한 인공신경망(920)은, 딥러닝을 수행하는 심층신경망 DNN 중 하나일 수 있다.
- [0254] 도 9의 (b)를 참조하면, 음성 감성 인식기(522)를 위한 인공신경망(920)은 사운드 데이터(925)가 입력되면, 입력된 사운드 데이터(925)의 특징점을 추출하고 사운드에 나타나는 감성을 인식할 수 있다.
- [0255] 음성 감성 인식기(522)를 위한 인공신경망(920)이 출력하는 감성 인식 결과(930)는 놀람, 행복, 슬픔, 싫음, 분노, 두려움, 중립 중 선택된 어느 하나의 감성 클래스일 수 있다. 더욱 바람직하게는 감성 인식 결과(830)는 놀람 x%, 행복 x%, 슬픔 x%, 싫음 x%, 분노 x%, 두려움 x%, 중립 x%와 같이 각 감성 클래스 별 확률값을 포함할 수 있다.
- [0256] 음성 감성 인식기(522)도 멀티모달 인식기(510)의 전처리기로 동작할 수 있다. 도 9의 (c)를 참조하면, 음성 감성 인식기(522)는 감성 인식 결과(930)를 출력할 뿐만 아니라, 입력된 사운드 데이터에 기초하여 추출된 특징점 벡터인 히든 스테이트(940)를 출력할 수 있다.
- [0257] 도 10은 본 발명의 실시예에 따른 멀티모달 감성 인식에 관한 설명에 참조되는 도면이고, 도 11은 본 발명의 일 실시예에 따른 감성 인식 결과를 예시한 도면이다.
- [0258] 도 10을 참조하면, 로봇(100) 또는 서버(70)가 구비하는 감성 인식기(74a)는 사용자가 발화한 음성의 내용을 포함하는 텍스트 유니모달 입력 데이터(1011), 사용자가 발화한 음성의 사운드를 포함하는 음성 유니모달 입력 데이터(1012), 사용자의 얼굴 이미지를 포함하는 영상 유니모달 입력 데이터(1013)가 입력될 수 있다.
- [0259] 또한, 감성 인식기(74a)에는 사용자를 포함하는 동영상 데이터가 수신되고, 모달 분리기(530)는 입력 데이터에 포함되는 오디오 데이터의 내용을 텍스트 데이터로 변환한 텍스트 유니모달 입력 데이터(1011)와 음성 톤, 크기, 높이 등 오디오 데이터의 음성(sound) 유니모달 입력 데이터(1012)로 분리하고, 동영상 데이터에서 사용자의 얼굴 이미지를 포함하는 영상 유니모달 입력 데이터(1013)를 추출할 수 있다.
- [0260] 한편, 유니모달 입력 데이터(1011, 1012, 1013)에 대한 전처리(Preprocess)가 수행될 수 있다.
- [0261] 예를 들어, 전처리 단계(1051, 1052, 1053)에서는 텍스트, 음성, 이미지 유니모달 입력 데이터(1011, 1012, 1013)에 포함된 노이즈를 제거하거나 감성 인식에 적합하도록 데이터를 추출, 변환하는 과정이 수행될 수 있다.
- [0262] 전처리가 끝나면, 유니모달 인식기들(521, 522, 523)은 각각 입력되는 유니모달 입력 데이터(1011, 1012, 1013)에서 감성을 인식하고 감성 인식 결과를 출력할 수 있다.
- [0263] 또한, 유니모달 인식기들(521, 522, 523)은 각각 입력되는 유니모달 입력 데이터(1011, 1012, 1013)에 기초하여 추출한 특징점 벡터를 멀티모달 인식기(510)로 출력할 수 있다.
- [0264] 멀티모달 인식기(510)의 병합기(512)는, 특징점 벡터들을 결합(Vector Concatenation)하여 멀티모달 감성 인식기(511)로 출력할 수 있다.
- [0265] 멀티모달 감성 인식기(511)는 3개의 유니모달 입력 데이터(1011, 1012, 1013)에 기초하는 멀티모달 입력 데이터에 대하여 감성 인식을 수행할 수 있다.
- [0266] 이를 위해, 멀티모달 감성 인식기(511)는 멀티모달 입력 데이터로 사전에 딥러닝된 인공신경망을 포함할 수 있다.
- [0267] 예를 들어, 멀티모달 감성 인식기(511)는 현재의 히든 스테이트가 직전의 히든 스테이트를 받아 갱신되는 순환 구조를 가지는 순환신경망(recurrent neural networks)을 포함할 수 있다. 한편, 멀티모달은 연관된 데이터들이

입력되므로 입출력이 독립적인 다른 인공신경망보다 순환신경망을 이용하는 것이 유리할 수 있다. 특히, 멀티모달 감성 인식기(511)는 순환신경망의 성능을 개선한 LSTM(long short term memory)을 포함할 수 있다.

- [0268] 로봇(100) 또는 서버(70)가 구비하는 감성 인식기(74a)는 복수의 딥러닝 구조를 가질 수 있다.
- [0269] 감성 인식기(74a)에서 3개의 유니모달 인식기들(521, 522, 523)과 1개의 멀티모달 감성 인식기(511)는 계층적 신경망(hierarchical neural network) 구조를 형성할 수 있다. 또한, 멀티모달 감성 인식기(511)는 순환신경망 이외의 인공신경망을 더 포함하여 자체적으로 계층적 신경망을 구성할 수도 있다.
- [0270] 한편, 감성 인식기(74a)는 감성 인식 결과(1090)를 출력한다.
- [0271] 감성 인식기(74a)는 유니모달(음성/영상/텍스트)과 멀티모달로 사용자의 감성을 7종의 감성 클래스 별 레벨(화률)로 인식할 수 있다.
- [0272] 감성 인식기(74a)는 입력으로 들어오는 사용자의 음성, 영상, 텍스트, 음성+영상+텍스트의 총 4가지 모달별로 감성을 인식할 수 있어, 사용자와의 정교한 인터랙션에 도움을 줄 수 있다.
- [0273] 유니모달(음성/영상/텍스트) 3개의 입력에 대한 출력은 각 입력별 감성 인식 값과 그 감성을 잘 표현하는 특징점 벡터일 수 있다.
- [0274] 감성을 잘 표현하는 특징점 벡터는 전결합 레이어(Fully Connected Layer)와 LSTM(long short term memory)을 이용하는 결합(Fusion) 기법 등으로 결합될 수 있다. 이에 따라 3가지 유니모달 입력이 결합되어 감성을 인식할 수 있다.
- [0275] 도 11은 인식된 감성 결과값의 일예를 도시한 것이다.
- [0276] 도 11을 참조하면, 음성, 영상, 텍스트의 각 유니모달 입력 데이터에 대한 유니모달 감성 인식 결과는 싫음, 중립, 행복으로 각각 출력될 수 있다.
- [0277] 또한, 음성, 영상, 텍스트의 특징점 벡터를 결합한 후에 감성 인식을 수행한 멀티모달 감성 인식 결과는 싫음 50%, 행복 43% 등 각 감성 클래스별 확률값으로 출력될 수 있다.
- [0278] 더욱 바람직하게는 유니모달 감성 인식 결과도 각 감성 클래스별 확률값으로 출력될 수 있다.
- [0279] 본 발명에 따른 감성 인식기(74a)는, 영상, 사운드뿐만 아니라 텍스트에 대한 정보까지 사용함으로써 인식 성능을 향상할 수 있다.
- [0280] 또한, 특정 유니모달 입력 데이터가 불충분하더라도 다른 유니모달 입력 데이터, 멀티모달 입력 데이터를 통해서 인식함으로써, 상호 보완적인 인식기(511, 521, 522, 523)를 구성할 수 있다.
- [0281] 또한, 총 4가지 정보의 조합으로 다양한 감성 인식이 가능하다.
- [0282] 도 11은 도 10에서 예시된 입력(1011, 1012, 1013)을 인식한 것으로, 사용자의 얼굴은 웃는 표정이나 사용자가 말한 내용에는 부정적 어휘가 사용된 경우의 감성 인식 결과이다.
- [0283] 이와 같이, 사람의 감정은 하나로 보기는 어려운 점이 많고, 표정과 말이 다른 모순되거나 복합적인 감성도 실제 생활 환경에서는 자주 발생할 수 있다.
- [0284] 현재 많은 연구에서 하나의 감성으로만 감성 인식 결과를 도출하고 있지만, 도 11과 같이 표정과 말이 다른 모순된 감성 상태 해당하는 경우엔 하나의 감성만 매핑할 경우 오인식의 가능성이 높아진다.
- [0285] 하지만, 본 발명에 따른 감성 인식 방법은 3가지의 입력에 대한 출력과 최종적으로 결합해서 나오는 출력값, 총 4가지의 감성 확률값을 통해 다양한 조합이 가능하다.
- [0286] 본 발명에 따른 감성 인식기(74a)는 영상, 음성, 문자 데이터를 상호보완적으로 이용하고 융합함으로써 사용자의 복잡한 감정 상태를 인식할 수 있다. 이에 따라, 감성 인식기(74a)는 실제 생활에서의 다양한 상황에 대해서 감성 인식이 가능하다.
- [0287] 또한, 본 발명에 따른 감성 인식기(74a)는 복잡한 감정 상태를 파악하기 때문에, 사용자의 심리 치료 로봇 등에 활용될 가능성이 높다. 예를 들어, 사용자로부터 부정적인 감정이 인지되더라도 감성 인식기(74a)를 포함하는 로봇(100)이 긍정적인 감성표현으로 감성케어(치료) 서비스를 제공할 수 있다.
- [0288] 도 12는 본 발명의 일 실시예에 따른 감성 인식 후처리에 관한 설명에 참조되는 도면이고, 도 13은 본 발명의

일 실시예에 따른 로봇의 감성 교류 사용자 경험에 관한 설명에 참조되는 도면이다.

- [0289] 도 12를 참조하면, 본 발명의 일 실시예에 따른 감성 인식기(74a)는, 복합 감성 인식 결과(1210)가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 소정 기준에 따라 최종 감성 인식 결과를 출력하는 후처리기(1220)를 더 포함할 수 있다.
- [0290] 본 발명의 일 실시예에 따른 로봇(100)은 후처리기(1220)를 포함할 수 있다.
- [0291] 본 발명의 일 실시예에 따른 로봇(100)은, 후처리기(1220)를 구비하는 감성 인식기(74a)를 포함하거나, 감성 인식기(74a)는 포함하지 않고 후처리기(1220)만을 포함할 수 있다.
- [0292] 설정에 따라, 후처리기(1220)는, 상기 복합 감성 인식 결과(1210)가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 모달별 인식기들(521, 522, 523)의 감성 인식 결과들 중에서 상기 멀티모달 인식기(511)의 감성 인식 결과와 일치하는 감성 인식 결과를 상기 최종 감성 인식 결과로 출력할 수 있다.
- [0293] 도 12의 예에서, 후처리기(1220)는, 텍스트 감성 인식기의 출력인 '싫음'과 상기 멀티모달 인식기(511)의 감성 인식 결과 중 확률값이 가장 높은 '싫음'이 일치하므로, 후처리기(1220)는, '싫음'을 최종 감성 인식 결과로 출력할 수 있다.
- [0294] 또는, 후처리기(1220)는, 상기 복합 감성 인식 결과(1210)가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 복합 감성 인식 결과(1210) 중 2개의 감성 클래스를 포함하는 모순 감성을 상기 최종 감성 인식 결과로 출력할 수 있다.
- [0295] 이 경우에, 후처리기(1220)는, 상기 멀티모달 인식기(511)의 감성 인식 결과 중 확률이 가장 높은 2개의 감성 클래스를 상기 모순 감성으로 선택할 수 있다.
- [0296] 도 12의 예에서 '싫음'과 '행복' 감성 클래스를 포함하는 모순 감성을 상기 최종 감성 인식 결과로 출력할 수 있다.
- [0297] 한편, 본 발명의 일 실시예에 따른 로봇(100)은 감성 인식기(74a)를 구비하여 사용자의 감성을 인식할 수 있다. 또는 본 발명의 일 실시예에 따른 로봇(100)은 감성 인식기(74a)를 구비하는 서버(70)와 통신하여 사용자의 감성 인식 결과를 수신할 수 있다.
- [0298] 본 발명의 일 실시예에 따른 로봇(100)은 사용자를 포함하는 동영상 데이터를 서버(70)로 송신하고, 상기 서버(70)로부터 상기 사용자의 감성 인식 결과를 복수개 포함하는 복합 감성 인식 결과를 수신하는 통신부(190) 및, 상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 일치하지 않는 2이상의 인식 결과를 조합하여 사용자의 감정을 확인하는 질문을 발화하는 음향 출력부(181)를 포함할 수 있다.
- [0299] 도 13에 나열된 다양한 예와 같이, 2이상의 다른 인식 결과를 포함하는 모순 감성에 해당하는 감성 인식 결과가 있다면, 로봇(100)은 사용자에게 모순 감성에 대한 질문을 발화할 수 있다.
- [0300] 예를 들어, 로봇(100)은, 상기 2개의 감성 클래스를 조합하여 사용자의 감정을 확인하는 질문을 발화할 수 있다.
- [0301] 또한, 로봇(100)은 상기 서버(70)로부터 수신된 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 소정 기준에 따라 최종 감성 인식 결과를 출력하는 후처리기(1220)를 더 포함할 수 있다.
- [0302] 상기 후처리기(1220)는, 상기 복합 감성 인식 결과가 일치하지 않는 2이상의 인식 결과를 포함하는 경우에, 상기 복합 감성 인식 결과 중 2개의 감성 클래스를 포함하는 모순 감성을 상기 최종 감성 인식 결과로 출력할 수 있다.
- [0303] 경우에 따라서, 상기 후처리기(1220)는, 상기 복합 감성 인식 결과 중 확률이 가장 높은 2개의 감성 클래스를 상기 모순 감성으로 선택할 수 있다.
- [0304] 본 발명에 따르면, 사용자는 로봇(100)의 질문에 대답하면서 로봇(100)과 교감할 수 있고, 자신의 감성을 이해하고 대화하는 로봇(100)에 대한 만족도가 증대될 수 있다.
- [0305] 또한, 사용자로부터 부정적인 감정이 인지되더라도 로봇(100)이 긍정적인 감성표현으로 감성케어(치료) 서비스를 제공할 수 있다.
- [0306] 실시예에 따라서, 사용자는 로봇(100)을 이용하여 화상 통화를 수행할 수 있고, 감성 인식기(74a)는 수신되는

화상 통화 데이터에 기초하여 화상 통화 상대방의 감성을 인식할 수 있다.

- [0307] 즉, 상기 감성 인식기(74a)는 화상 통화 상대방의 화상 통화 데이터를 입력받고 상기 화상 통화 상대방의 감성 인식 결과를 출력할 수 있다.
- [0308] 또한, 감성 인식은 감성 인식기(74a)를 구비하는 서버(70)에서 수행될 수 있다. 예를 들어, 사용자는 로봇(100)을 이용하여 화상 통화를 수행할 수 있고, 서버(70)는 로봇(100)으로부터 화상 통화 데이터를 수신하고 상기 수신된 화상 통화 데이터에 포함된 사용자의 감성 인식 결과를 로봇(100)으로 송신할 수 있다.
- [0309] 도 14는 본 발명의 일 실시예에 따른 감성 인식기의 동작 방법을 도시한 순서도이다.
- [0310] 도 14를 참조하면, 본 발명의 일 실시예에 따른 감성 인식기(74)는 데이터가 입력되면(S1410), 입력 데이터에 기초하여 복수의 유니모달 입력 데이터를 생성할 수 있다(S1420). 예를 들어, 모달 분리기(530)는 입력 데이터를 분리하여 복수의 유니모달 입력 데이터를 생성할 수 있다(S1420).
- [0311] 각각의 모달별 인식기(521, 522, 523)는 대응하는 유니모달 입력 데이터에서 사용자의 감성을 인식할 수 있다(S1430).
- [0312] 모달별 인식기(521, 522, 523)는 감성 인식 결과와 유니모달 입력 데이터의 특징점 벡터를 출력할 수 있다.
- [0313] 모달별 인식기(521, 522, 523)가 출력한 특징점 벡터들은 병합기(512)에서 병합되고, 멀티 모달 감성 인식기(511)는 병합된 멀티모달 데이터에 대해서 감성 인식을 수행할 수 있다(S1450).
- [0314] 한편, 본 발명의 일 실시예에 따른 감성 인식기(74)는 모달별 인식기(521, 522, 523)의 감성 인식 결과들과 멀티 모달 감성 인식기(511)의 감성 인식 결과를 포함하는 복합 감성 인식 결과를 출력할 수 있다.
- [0315] 이와 같이, 본 발명의 일 실시예에 따른 감성 인식기(74)는 유니모달과 멀티모달의 4개의 딥러닝 기반 감성인식 모델을 구성하여, 한 시점에 입력되는 사용자의 감성을 유니모달 별로 인식하면서 멀티모달로도 복합 인식할 수 있다.
- [0316] 또한, 감성 인식기(74)는 유니모달 감성 인식 결과와 멀티모달 감성 인식 결과를 감성 클래스별 레벨(확률)로 출력할 수 있다.
- [0317] 이에 따라, 음성, 영상, 텍스트 각각의 감성을 인식하여 사용자만의 특정한 감성 피드백이 가능하고, 음성, 영상, 텍스트를 종합한 멀티모달 감성 인식 결과도 종합적으로 이용할 수 있다.
- [0318] 본 발명의 실시예들 중 적어도 하나에 의하면, 사용자 감성(emotion)을 인식하고 감성기반의 서비스를 제공할 수 있다.
- [0319] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 딥러닝으로 학습된 인공지능을 이용하여 사용자의 감성을 더욱 정확하게 인식할 수 있다.
- [0320] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복수의 감성 인식 결과를 출력하고, 감성 인식 결과들을 조합하여 다양한 방식으로 이용할 수 있다.
- [0321] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복수의 감성 인식 결과에 기초하여 사용자와 대화함으로써 사용자와 교감하며 더욱 정확하게 사용자의 감성을 인식할 수 있다.
- [0322] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 니모달과 멀티모달 감성 인식을 개별적으로 수행하고 복수의 인식 결과를 상호보완적으로 이용함으로써, 사용자의 감성을 더욱 정확하게 인식할 수 있다.
- [0323] 또한, 본 발명의 실시예들 중 적어도 하나에 의하면, 복합적인 감성을 인식할 수 있어, 사용자의 만족도와 편의성을 높일 수 있다.
- [0324] 본 발명에 따른 감성 인식기, 감성 인식기를 포함하는 로봇 및 서버는 상기한 바와 같이 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.
- [0325] 한편, 본 발명의 실시예에 따른 감성 인식기, 로봇 및 서버의 동작 방법은, 프로세서가 읽을 수 있는 기록매체에 프로세서가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 프로세서가 읽을 수 있는 기록매체는 프로세서에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 프로세서가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한, 인터넷을

통한 전송 등과 같은 캐리어 웨이브의 형태로 구현되는 것도 포함한다. 또한, 프로세서가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 프로세서가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0326] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안될 것이다.

부호의 설명

[0328] 홈 어플라이언스: 10

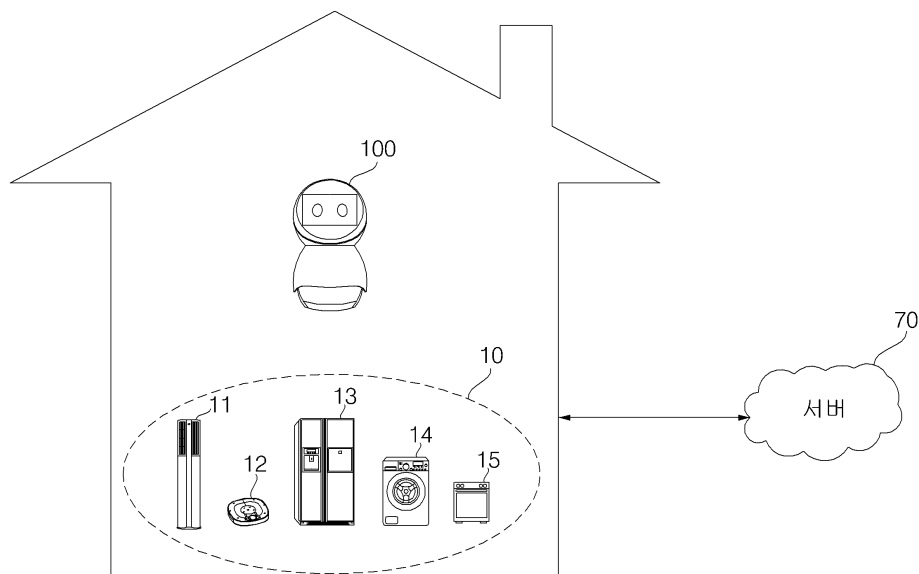
서버: 70

로봇: 100

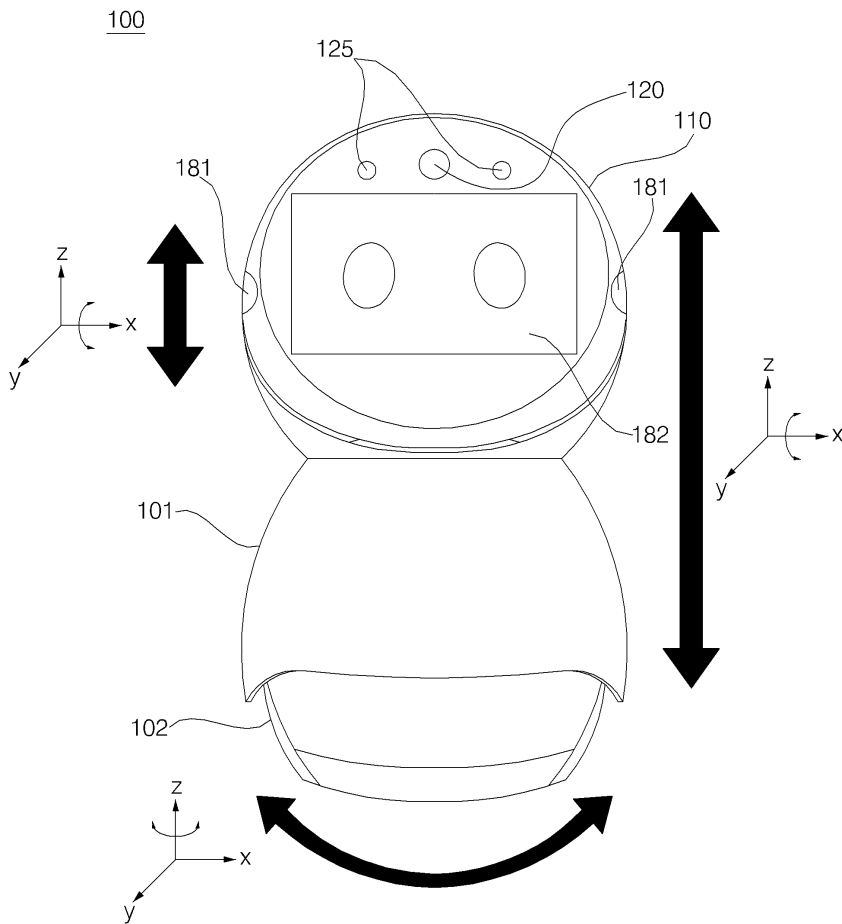
감성 인식기: 74a

도면

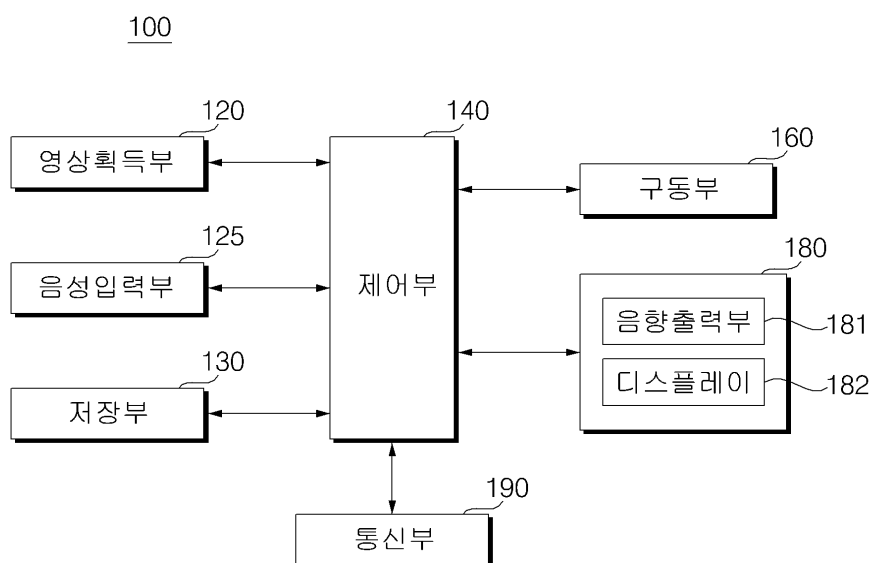
도면1



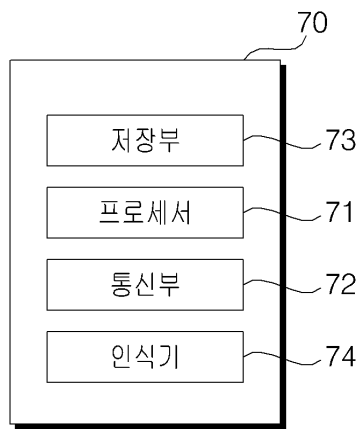
도면2



도면3

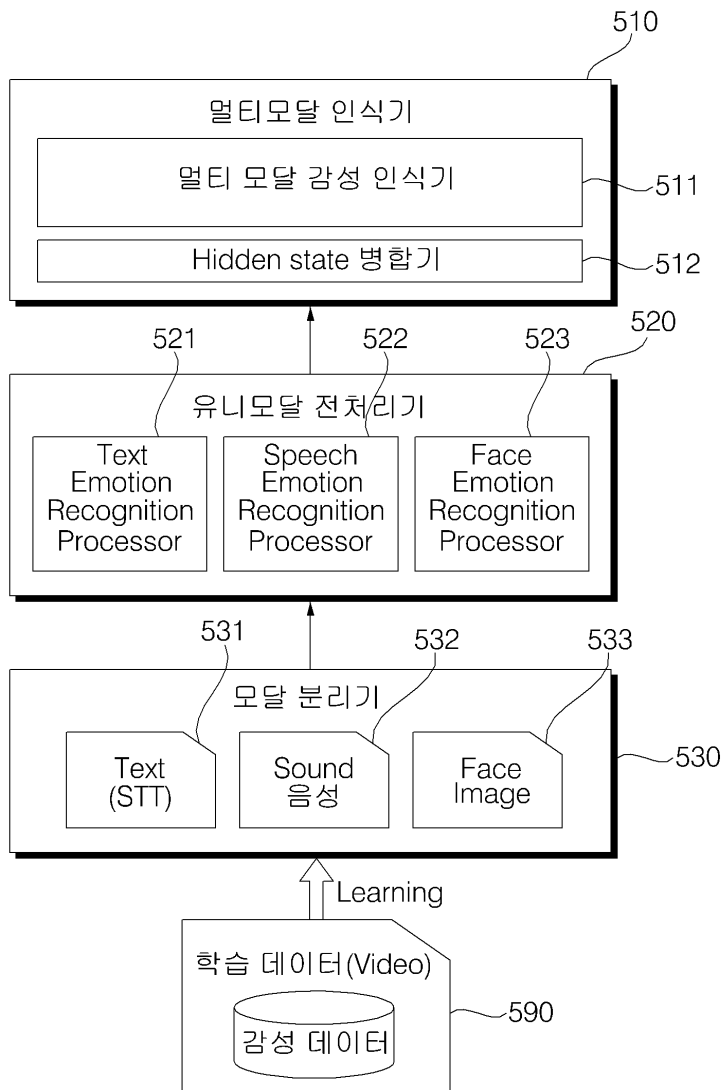


도면4

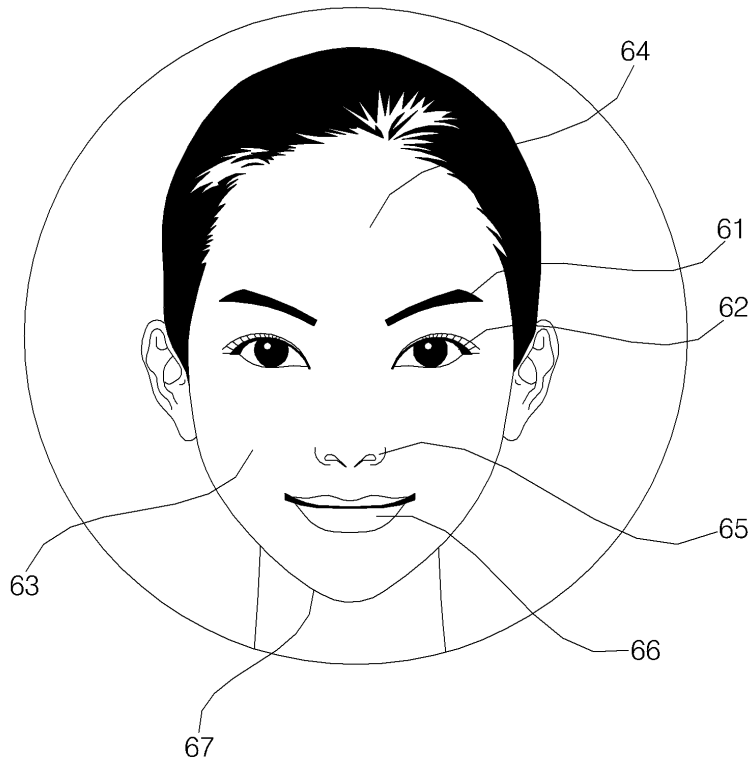


도면5

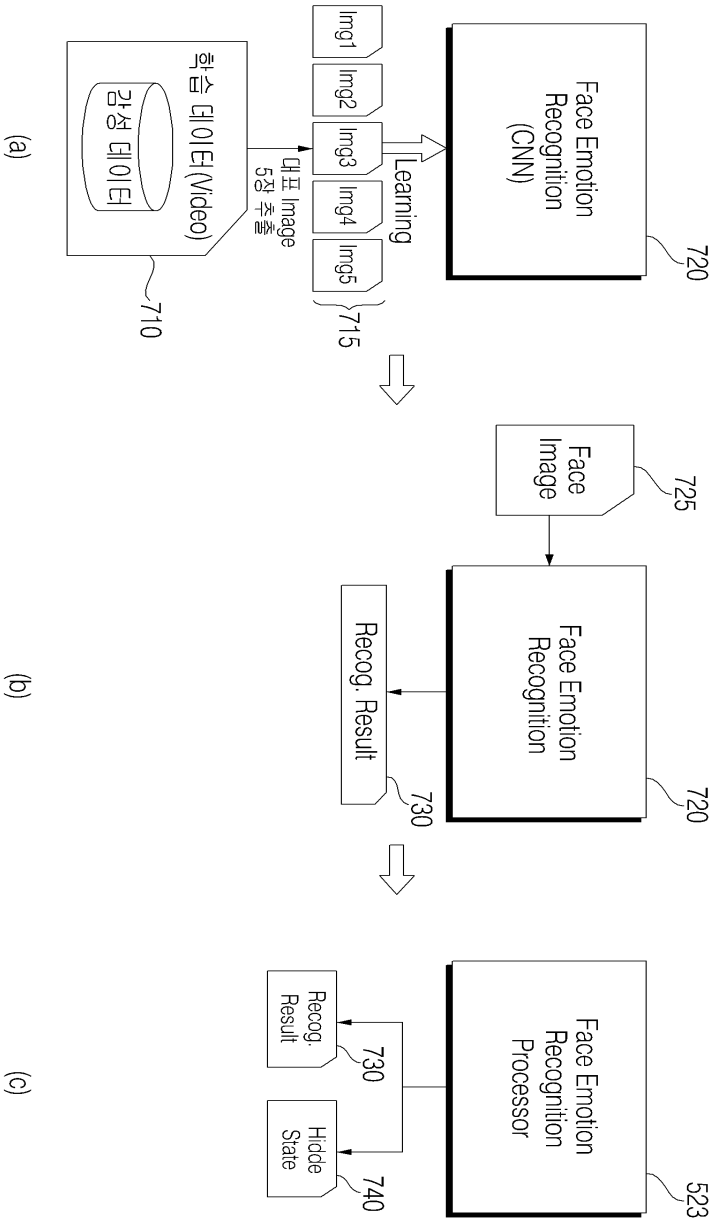
74a



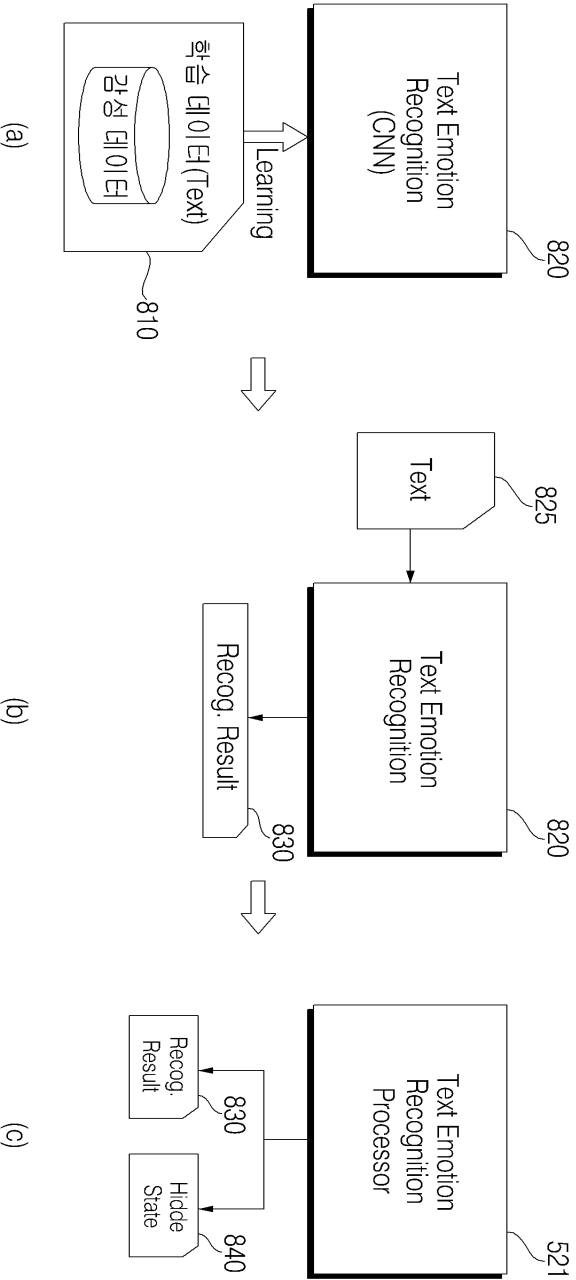
도면6



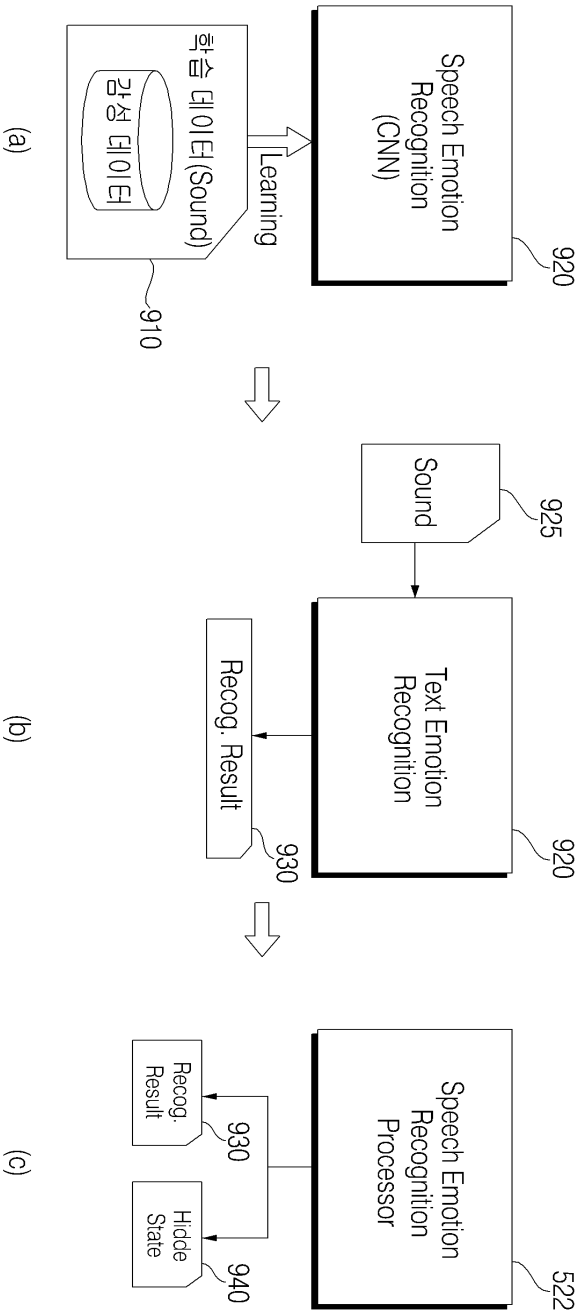
도면7



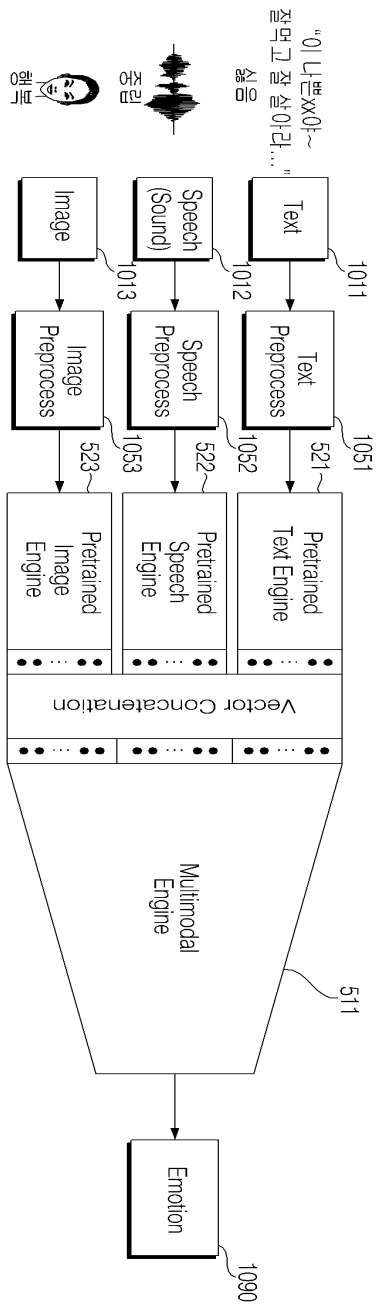
도면8



도면9



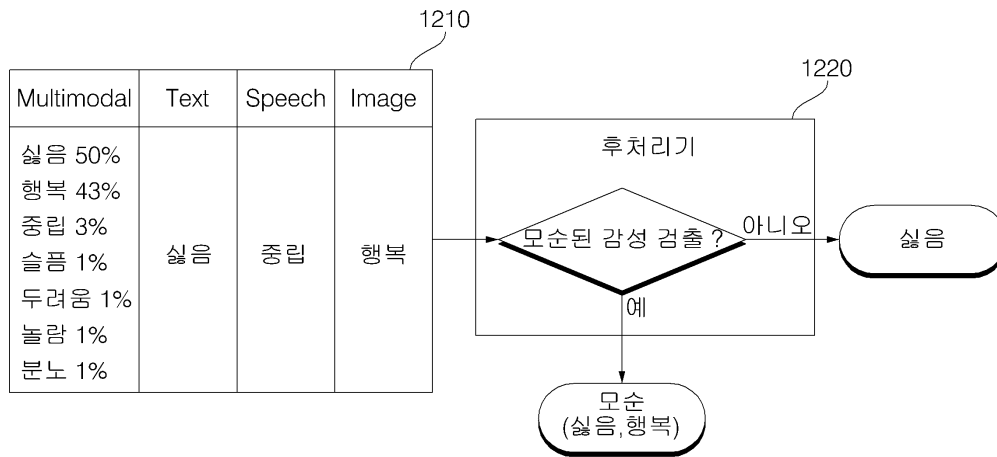
도면10



도면11

Multimodal	Text	Speech	Image
싫음 50% 행복 43% 중립 3% 슬픔 1% 두려움 1% 놀람 1% 분노 1%	싫음	중립	행복

도면12



도면13

모순감성 Index	Multimodal	Text	Speech	Image	감성교류 UX
1	싫음 53%, 행복 43%	싫음 90%	중립 80%	행복 89%	"~~님, 진심이세요? 얼굴은 행복해보이시는데"
2	행복 50% 슬픔 42%	행복 95%	행복 94%	슬픔 92%	"행복의 눈물을 흘리시네요. ~~님이 행복해 하시니 저도 눈물이 나네요"
3	싫음 52% 놀람 46%	행복 99%	싫음 77%	놀람 99%	"사실 싫으시죠?" - Scene: 남이 잘못을 때 배아플때
4	놀람 55% 중립 40%	놀람 99%	중립 89%	놀람 66%	"에이 놀란척 하지 마세요."
5	중립 60% 행복 38%	중립 97%	행복 88%	행복 68%	"티 안내는 척 하지만 기쁜건가요?"
6	행복 56% 싫음 42%	행복 99%	싫음 79%	중립 93%	"솔직하게 말해봐요 머가 싫은지" - Scene: 선물 받았지만 마음에 안들 때
7	중립 54% 두려움 44%	중립 88%	두려움 99%	두려움 88%	"혹시 떨고 계신가요?" - Scene: 남자가 허세 부릴 때
8	두려움 51% 중립 47%	두려움 99%	중립 88%	두려움 86%	"무서운거 맞아요? 아닌것 같은데" - Scene: 여자친구가 무서운 척 할 때

도면14

