



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년04월25일

(11) 등록번호 10-2390781

(24) 등록일자 2022년04월21일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2022.01) **A63F 13/212** (2014.01)
A63F 13/327 (2014.01) **A63F 13/56** (2014.01)
G02B 27/00 (2020.01) **G02B 27/01** (2006.01)
G06F 3/01 (2006.01) **H04N 5/225** (2006.01)
H04N 7/18 (2006.01)
(52) CPC특허분류
G06V 40/176 (2022.01)
A63F 13/212 (2015.01)
(21) 출원번호 10-2016-7032126
(22) 출원일자(국제) 2015년04월23일
심사청구일자 2020년03월25일
(85) 번역문제출일자 2016년11월17일
(65) 공개번호 10-2016-0146861
(43) 공개일자 2016년12월21일
(86) 국제출원번호 PCT/US2015/027188
(87) 국제공개번호 WO 2015/167909
국제공개일자 2015년11월05일
(30) 우선권주장
14/265,201 2014년04월29일 미국(US)
(56) 선행기술조사문헌
KR101133225 B1*
KR1020120018479 A*
WO2013077076 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
장 차
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠
(8/1172) 마이크로소프트 테크놀로지 라이선싱,
엘엘씨 내
장 쟁유
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠
(8/1172) 마이크로소프트 테크놀로지 라이선싱,
엘엘씨 내
로메라 파레데스 베르나르디노
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠
(8/1172) 마이크로소프트 테크놀로지 라이선싱,
엘엘씨 내
(74) 대리인
김태홍, 김진희

전체 청구항 수 : 총 20 항

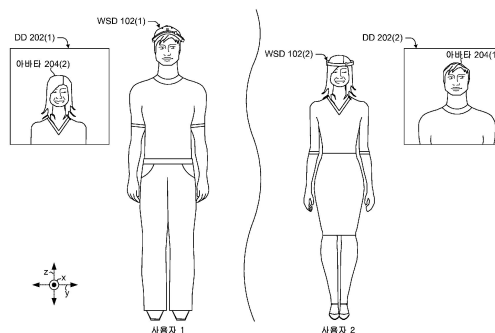
심사관 : 노용완

(54) 발명의 명칭 **얼굴 표정 추적**

(57) 요약

본 설명은 얼굴 추적에 관한 것이다. 일례는 사용자의 얼굴에 대해 웨어러블 디바이스를 배치하도록 구성된 배향 구조(orientation structure)를 포함할 수 있다. 이 예는 또한 사용자의 얼굴에 걸쳐 이미지를 캡처하기 위해 배향 구조에 의해 사용자의 얼굴과 평행하거나 사용자의 얼굴과의 로우 앵글(low angle)로 고정된 카메라를 포함할 (뒷면에 계속)

대표도



수 있다. 이 예는 이미지를 수신하고 그 이미지를 아바타 모델과 연관된 파라미터에 매핑하도록 구성된 프로세서를 더 포함할 수 있다.

(52) CPC특허분류

A63F 13/327 (2015.01)

A63F 13/56 (2015.01)

G02B 27/0093 (2013.01)

G02B 27/0172 (2013.01)

G06F 3/011 (2022.02)

G06V 40/19 (2022.01)

H04N 5/2252 (2013.01)

H04N 7/18 (2013.01)

G02B 2027/0178 (2013.01)

명세서

청구범위

청구항 1

시스템에 있어서,

스마트 안경(smart eyeglasses)과,

프로세서

를 포함하고,

상기 스마트 안경은,

상기 스마트 안경을 착용한 사용자가 바라보게 배치된 한 쌍의 렌즈를 고정시키는 프레임과,

상기 프레임에 대해 고정되며, 상기 스마트 안경을 착용한 사용자의 뺨 또는 턱의 이미지를 포함하는 로우 앵글(low angle) 센서 데이터를 캡처하도록 구성된 적어도 하나의 센서

를 포함하고, 상기 적어도 하나의 센서는 상기 스마트 안경을 착용한 사용자의 얼굴에 평행한 기준면에 대해 45도 미만의 각도로 배향되며,

상기 프로세서는,

상기 로우 앵글 센서 데이터를 취득하고,

상기 로우 앵글 센서 데이터로부터 상기 뺨 또는 턱의 이미지를, 상기 스마트 안경을 착용한 사용자를 표현하는 아바타 모델과 연관된 입 움직임(mouth movement)을 포함한 애니메이션 파라미터에 상관시키기 위한 매핑(mapping) 및 상기 아바타 모델을 사용하며,

상기 애니메이션 파라미터를 출력하도록

구성되고,

상기 매핑은, 트레이닝 스테이지 동안, 상기 사용자의 전체 얼굴 이미지를 캡처하도록 상기 사용자의 정면에 배치된 기준 카메라 및 상기 적어도 하나의 센서에 의해 동시에 캡처된 사용자 액션에 기초하여, 특정 사용자 액션을 특정 애니메이션 파라미터에 관련시키도록 생성되는 것인, 시스템.

청구항 2

제1항에 있어서, 상기 스마트 안경은,

상기 스마트 안경을 착용한 사용자의 눈을 향하도록 상기 프레임 상에 배향되며, 상기 사용자의 눈의 다른 이미지를 캡처하도록 구성된 적어도 하나의 다른 센서를 더 포함하고,

상기 프로세서는, 상기 아바타 모델을 사용하여 상기 이미지와 상기 다른 이미지 둘 다를 처리함으로써 상기 애니메이션 파라미터를 취득하도록 구성되는 것인, 시스템.

청구항 3

제2항에 있어서, 상기 적어도 하나의 센서와 상기 적어도 하나의 다른 센서는 가시광 카메라 또는 적외선 카메라를 포함하는 것인, 시스템.

청구항 4

제1항에 있어서, 상기 적어도 하나의 센서는 개별 이미지에서 상기 스마트 안경을 착용한 사용자의 입과 상기 사용자의 양쪽 눈 중 적어도 한쪽 눈을 캡처할 수 있는 어안 렌즈를 포함하는 것인, 시스템.

청구항 5

제1항에 있어서, 상기 프레임은 한 쌍의 안경다리를 포함하고, 상기 적어도 하나의 센서는 상기 안경다리 쌍 중 개별 안경다리의 장축에 수직으로 배향되는 것인, 시스템.

청구항 6

제1항에 있어서,

상기 스마트 안경과 분리된 컴퓨팅 디바이스를 더 포함하고, 상기 프로세서는 상기 컴퓨팅 디바이스에 내장되며,

상기 스마트 안경은, 상기 로우 앵글 센서 데이터를 상기 컴퓨팅 디바이스에 전송하도록 구성된 회로부를 더 포함하는 것인, 시스템.

청구항 7

제1항에 있어서, 상기 프로세서는 상기 스마트 안경 상에 제공되는 것인, 시스템.

청구항 8

제1항에 있어서, 상기 입 움직임을 결정하는데 사용되는 상기 뺨 또는 턱의 이미지는, 상기 스마트 안경을 착용한 사용자의 입의 이미지를 포함하지 않는 것인, 시스템.

청구항 9

웨어러블(wearable) 스마트 디바이스에 있어서,

사용자의 얼굴에 대해 상기 웨어러블 스마트 디바이스를 위치 결정하도록 구성된 배향 구조(orientation structure)와,

상기 사용자의 입과 상기 사용자의 눈의 정면을 포함하거나 통과하는 기준면에 대해 ± 45 도 미만의 각도 또는 평행한 방향으로 상기 사용자의 입의 이미지를 캡처하도록, 상기 배향 구조에 의해 고정된 카메라와,

프로세서

를 포함하고,

상기 프로세서는 상기 이미지를 수신하고, 아바타 모델과 연관된 입 움직임을 포함하는 애니메이션 파라미터에 상기 사용자의 입의 이미지를 매핑하도록 구성 - 상기 매핑은, 트레이닝 스테이지 동안, 상기 카메라에 의해, 그리고 상기 사용자의 전체 얼굴 이미지를 캡처하도록 상기 사용자의 정면에 배치된 기준 카메라에 의해 동시에 캡처된 사용자 액션에 기초하여, 특정 사용자 액션을 특정 애니메이션 파라미터에 관련시키도록 생성됨 - 되는 것인, 웨어러블 스마트 디바이스.

청구항 10

제9항에 있어서, 상기 웨어러블 스마트 디바이스는 스마트 안경이고, 상기 배향 구조는 프레임을 포함하는 것인, 웨어러블 스마트 디바이스.

청구항 11

제10항에 있어서, 상기 프레임에 의해 지지되는 한 쌍의 렌즈를 더 포함하고,

상기 스마트 안경은 상기 렌즈 쌍의 렌즈 중 한쪽 또는 양쪽에 이미지를 제시하도록 구성되는 것인, 웨어러블 스마트 디바이스.

청구항 12

제11항에 있어서, 상기 렌즈 쌍은 교정 렌즈인 것인, 웨어러블 스마트 디바이스.

청구항 13

제9항에 있어서, 상기 웨어러블 스마트 디바이스는 스마트 햇(smart hat)이고, 상기 배향 구조는 상기 스마트 햇의 뺨(bill)을 포함하는 것인, 웨어러블 스마트 디바이스.

청구항 14

제9항에 있어서, 상기 배향 구조는 헤드 밴드(head band)를 포함하는 것인, 웨어러블 스마트 디바이스.

청구항 15

제9항에 있어서, 상기 배향 구조는 프레임을 포함하고, 상기 웨어러블 스마트 디바이스는 상기 사용자의 의복에 상기 프레임을 고정시키는 클립을 더 포함하는 것인, 웨어러블 스마트 디바이스.

청구항 16

제10항에 있어서, 상기 사용자의 눈을 향하도록 상기 프레임 상에 배향되며, 상기 카메라가 상기 사용자의 입의 이미지를 캡처하는 것과 동시에 상기 사용자의 눈의 다른 이미지를 캡처하도록 구성된 적어도 하나의 상이한 센서를 더 포함하고, 상기 프로세서는 상기 이미지 및 상기 다른 이미지로부터 얼굴 파라미터를 취득하도록 구성되는 것인, 웨어러블 스마트 디바이스.

청구항 17

제9항에 있어서, 상기 카메라는 개별 이미지에서 상기 사용자의 입과 상기 사용자의 양쪽 눈 중 적어도 한쪽 눈을 캡처할 수 있는 어안 렌즈를 포함하는 것인, 웨어러블 스마트 디바이스.

청구항 18

제10항에 있어서, 상기 프레임은 한 쌍의 안경다리를 포함하고, 상기 카메라는 상기 안경다리 쌍 중 개별 안경다리의 장축에 수직으로 배향되는 것인, 웨어러블 스마트 디바이스.

청구항 19

컴퓨터 구현 방법(computer implemented method)에 있어서,

사용자의 전방에 배치된 기준 카메라로부터 정면 트레이닝 이미지를 취득하는 단계와,

상기 사용자의 얼굴에 대해 로우 앵글로 배치된 로우 앵글 카메라로부터 로우 앵글 트레이닝 이미지를 취득하는 단계로서, 상기 로우 앵글 트레이닝 이미지는 상기 사용자의 뺨 또는 턱을 캡처한 것인, 상기 로우 앵글 트레이닝 이미지를 취득하는 단계와,

상기 정면 트레이닝 이미지 및 상기 로우 앵글 트레이닝 이미지를 사용하여 모델을 트레이닝하는 단계와,

웨어러블 디바이스를 착용하고 있는 상기 사용자의 얼굴을 따라 로우 앵글로 캡처되는 로우 앵글 이미지 데이터를 수신하는 단계로서, 상기 로우 앵글 이미지 데이터는 상기 웨어러블 디바이스를 착용한 사용자의 상기 뺨 또는 턱의 이미지를 포함하는 것인, 상기 로우 앵글 이미지 데이터를 수신하는 단계와,

상기 뺨 또는 턱의 이미지를 입 애니메이션 파라미터에 매핑하도록 구성된 상기 모델로 상기 사용자의 뺨 또는 턱의 이미지를 프로세싱하는 단계와,

상기 사용자의 애니메이션화된 표현으로 하여금 상기 입 애니메이션 파라미터에 따라 입 움직임을 수행하게 하는 단계

를 포함하는, 컴퓨터 구현 방법.

청구항 20

제19항에 있어서, 상기 로우 앵글 이미지 데이터를 수신하는 단계는 네트워크를 통해 상기 웨어러블 디바이스로부터 상기 로우 앵글 이미지 데이터를 수신하는 단계를 포함하고, 상기 웨어러블 디바이스는 상기 로우 앵글 이미지 데이터를 캡처한 센서를 구비하는 것인, 컴퓨터 구현 방법.

청구항 21

삭제

발명의 설명

기술 분야

배경 기술

- [0001] 다양한 웨어러블 디바이스(wearable device)가 특정 용도로 출시되고 있다. 이를테면, 스마트 안경(smark glasses)과 헤드 마운티드 디스플레이(head-mounted display)는 사용자가 바라보고 있는 것을 결정할 수 있다. 그러나, 이들 디바이스는 사용자와 마주하여 배치되고 외향 카메라(outward pointing camera) 및/또는 시선 추적을 위해 사용자의 눈을 향하는 카메라를 구비하기 때문에 사용자의 유용한 뷰(view)를 제공하지 못한다. 본 개념은 카메라가 사용자에게 매우 근접하게 배치되어 있더라도 사용자의 유용한 뷰를 제공하여 웨어러블 디바이스에 대해 추가 기능을 제공할 수 있다.

발명의 내용

- [0002] 본 설명은 얼굴 추적에 관한 것이다. 일례는 사용자의 얼굴에 대해 웨어러블 디바이스를 배치하도록 구성된 배향 구조(orientation structure)를 포함할 수 있다. 이 예는 또한 사용자의 얼굴에 걸쳐 이미지를 캡처하기 위해 배향 구조에 의해 사용자의 얼굴과 평행하거나 사용자의 얼굴과의 로우 앵글(low angle)로 고정된 카메라를 포함할 수 있다. 이 예는 이미지를 수신하고 그 이미지를 아바타 모델과 연관된 파라미터에 매핑하도록 구성된 프로세서를 더 포함할 수 있다.
- [0003] 앞에 열거한 예는 본 명세서를 읽는 사람들을 돕도록 신속한 참고를 제공하는 것을 목적으로 하며, 본 명세서에 기술하는 원리의 범위를 한정하는 것을 의도하지 않는다.

도면의 간단한 설명

- [0004] 첨부하는 도면은 본 명세서에 포함된 개념의 구현을 예시한다. 예시하는 구현의 특징은 첨부 도면과 함께 이하의 설명을 참조함으로써 더욱 쉽게 이해될 수 있다. 여러 도면에서의 같은 참조 번호는 같은 엘리먼트를 가리키기 위해 가능한 곳이라면 어디에서나 사용된다. 또한, 각 참조 번호의 최좌측 숫자는 이 참조 번호가 맨처음 소개된 도면과 관련 설명을 나타낸다.
- 도 1, 도 4와 도 5는 본 개념의 일부 구현예에 따른 웨어러블 스마트 디바이스의 예를 도시하고 있다.
- 도 2와 도 3은 본 개념의 일부 구현예에 따른 시각화 시나리오의 예를 도시하고 있다.
- 도 6은 본 개념의 일부 구현예에 따른 예시적인 시각화 프로세스이다.
- 도 7 내지 도 11b는 본 개념의 일부 구현예에 따른 예시적인 시각화 시스템의 양태를 보다 상세하게 도시하고 있다.
- 도 12는 본 개념의 일부 구현예에 따른 예시적인 시각화 프로세스를 도시하고 있다.

발명을 실시하기 위한 구체적인 내용

- [0005] 개관
- [0006] 본 설명은 사용자에게 매우 근접하게 그리고/또는 로우 앵글로 배치된 카메라로부터의 얼굴 정보 등의 사용자에게 관한 유용한 시각적 정보를 캡처하는 것에 관한 것이다. 이 개념은 로우 앵글 카메라를 채택하는 다양한 웨어러블 스마트 디바이스에 대해 구현될 수 있다.
- [0007] 본 개념의 일 양태는 사용자의 로우 앵글 이미지로부터 가치있는 정보를 도출할 수 있는 특징을 포함한다. 이를테면, 카메라 또는 카메라들이 사용자에게 근접 배치될 수 있고 사용자의 얼굴에 거의 평행한 이미지를 캡처할 수 있다. 통상 이러한 이미지는 가치가 거의 없다. 본 구현예는 이들 이미지로부터 유용한 정보를 도출할 수 있다. 유용한 정보는 다양한 용도로, 예컨대 사용자의 아바타를 실물과 똑같은 방식으로 제어하는 데에 이용될 수 있다.
- [0008] 소개를 위한 도 1은 몇몇 웨어러블 스마트 디바이스(102)의 예를 도시하고 있다. 이 경우에, 웨어러블 스마트 디바이스(102(1))는 스마트 햇(smart hat)이고, 웨어러블 스마트 디바이스(102(2))는 스마트 헤드 밴드(smart head band)이며, 웨어러블 스마트 디바이스(102(3))는 스마트 안경이고, 웨어러블 스마트 디바이스(102(4))는

스마트 마이크(smart microphone)임이 자명하다. 물론, 다른 예도 고려된다. 웨어러블 스마트 디바이스(102)는 하나 이상의 센서를 포함할 수 있다. 이 예에서는, 이들 센서가 로우 앵글 카메라(104), 시선 추적 카메라(106), 및/또는 마이크(108)임이 자명하다. (웨어러블 스마트 디바이스(102(1), 102(3))는 로우 앵글 카메라와 시선 추적 카메라 둘다를 포함하는 것을 주목해야 한다. 반면, 웨어러블 스마트 디바이스(102(2), 102(4))는 시선 추적 카메라를 포함하지 않는다.) 카메라는 무엇보다도, 가시광 카메라 및/또는 적외선 카메라, 및/또는 깊이 카메라일 수 있다. 이들 엘리먼트의 기능에 대해서는 후술한다. 웨어러블 스마트 디바이스는 배향 구조(orientation structure)(110)도 포함한다. 배향 구조는 프레임, 밴드(band), 쉘(shell), 또는 사용자에게 대해 센서의 위치 및/또는 배향을 유지하는 것을 돕는 기타 구조임이 자명하다. 전술한 바와 같이, 예시하는 구현예 중 일부는 로우 앵글 카메라(104)와 시선 추적 카메라(106) 둘다를 사용한다. 다른 구현예는 이 2개의 카메라에 의해 다른 식으로 제공되는 데이터를 캡처할 수 있는 단일 카메라를 채택할 수 있다. 이를테면, 어안 렌즈를 구비한 로우 앵글 카메라(104)는 눈, 입, 뺨, 및/또는 턱 등의 사용자의 얼굴의 일부를 캡처하는데 사용될 수 있다.

[0009] 정리하면, 웨어러블 스마트 디바이스가 점점 흔해지고 있다. 웨어러블 스마트 디바이스는 종종 카메라를 포함한다. 그러나, 카메라가 사용자에게 대해 너무 근접해 있고/있거나 비일반적인 배향으로 배치되기 때문에, 카메라는 사용자, 특히 사용자의 얼굴의 유용한 이미지를 제공할 수 없다. 본 구현예는 이들 카메라를 배치하고, 사용자의 얼굴 표정, 입 움직임, 눈 깜박임, 시선 방향 등에 관해 유용한 정보를 생성하는 식으로 이미지를 프로세싱할 수 있다.

[0010] 도 2는 '사용자 1'이 착용하는 웨어러블 스마트 디바이스(102(1))와 '사용자 2'가 착용하는 웨어러블 스마트 디바이스(102(2))와 2개의 디스플레이 디바이스(202(1), 202(2))를 포함하는 사용례 시나리오를 나타내고 있다. 사용자 1은 디스플레이 디바이스(202(1))에 근접한 제1 위치에 있고, 사용자 2는 제1 위치로부터 약간의 거리를 둔 제2 위치에 있으며, 사용자 2는 디스플레이 디바이스(202(2))에 근접해 있다고 상정한다. 이 예에서는, (도 1에 도시하는 바와 같이), 웨어러블 스마트 디바이스(102(1))가 로우 앵글 카메라(104(1))와, 사용자 1의 눈에 지향된 다른 카메라(예컨대, 시선 추적 카메라)(106(1))를 포함한다. 웨어러블 스마트 디바이스(102(2))는 로우 앵글 카메라(104(2))를 포함한다.

[0011] 이 시나리오에서는, 웨어러블 스마트 디바이스(102(1))가 사용자 1의 이미지를 캡처한다. 이들 이미지는 사용자 1의 표현(representation)을 제어하는데 사용될 수 있다. 이 경우에, 사용자 1의 표현은 사용자 1에 근접한 디스플레이 디바이스(202(2)) 상에 제시되는 사용자 1의 아바타(204(1))임이 자명하다. 마찬가지로, 웨어러블 스마트 디바이스(102(2))는 사용자 2의 이미지를 캡처한다. 이들 이미지는 사용자 2의 표현을 제어하는데 사용될 수 있다. 이 경우에, 사용자 2의 표현은 사용자 1에 근접한 디스플레이 디바이스(202(1)) 상에 제시되는 사용자 2의 아바타(204(2))임이 자명하다. 이 시나리오에서는, 사용자들이 서로 통신할 수 있고, 이들 각각의 아바타는 다른 사용자에게 대해 이들의 눈 움직임, 얼굴 표정, 턱 움직임, 및/또는 (예컨대, 발음으로 인한) 입 움직임을 나타낼 수 있다.

[0012] 아바타(204)는, 눈 움직임, 얼굴 표정, 및/또는 입 움직임을 캡처하지만 사용자와 반드시 닮을 필요는 없는, 만화 캐릭터 또는 컴퓨터로 생성된 캐릭터 등의 임의의 형태로 표시될 수 있다. 다른 구현예에서는, 아바타가 사용자의 실제 표현(예컨대, 증강된 사진과 동일한(photo-realistic) 이미지)일 수도 있다. 토론중에 각각의 아바타를 볼 수 있는 특징은 음성만의 통신보다 더욱 생생하고(robust) 즐거운 토론을 가능하게 할 수 있다. 이를테면, 토론중에 다른 사용자의 반응을 볼 수 있는 특징은 얼굴을 마주한 대화를 모방하는 피드백을 제공할 수 있다.

[0013] 또한 도 2의 시나리오에는 2명의 사용자만이 포함되지만, 다른 구현예들은 추가 사용자들을 포함할 수도 있음을 주목해야 한다. 또한, 로우 앵글 카메라(104) 및/또는 시선 추적 카메라(106)가 사용자의 눈을 추적할 수 있기 때문에, 구현예들은 사용자들이 바라보고 있는 사용자 개인을 구별할 수 있다(예컨대, 사용자 1과 사용자 2가 사용자 3의 아바타를 바라보고 있다). 이 정보는 사용자의 아바타에 반영될 수 있고, 꼭 얼굴을 맞댄 그룹의 시나리오와 같이 통신 프로세스의 유용한 일부일 수 있다.

[0014] 도 3은 도 2의 사용례 시나리오와 유사한 다른 사용례 시나리오를 도시하고 있다. 이 특정 사용례 시나리오는 웨어러블 스마트 디바이스의 2 세트(102(3)(A), 102(3)(B))를 포함한다. 명백하게 나타내지는 않지만, 웨어러블 스마트 디바이스는 도 2의 사용자 1과 사용자 2가 착용한 것이다. 도 2에 대해 전술한 바와 유사하게, 웨어러블 스마트 디바이스(102(3)(A))는 사용자 1의 로우 앵글 이미지를 캡처하고, 이 이미지를, 사용자 2에게 표시될 수 있는 사용자 1의 표현(예컨대, 아바타(204(1)))을 제어하는데 이용할 수 있다. 마찬가지로, 웨어러블 스마트 디

바이스(102(3)(B))는 사용자 1에게 표시될 수 있는 사용자 2의 표현(예컨대, 아바타(204(2)))를 제어하기 위해 사용자 2에 대해 동일한 기능을 수행할 수 있다.

[0015] 이에, 웨어러블 스마트 디바이스(102(3)(A))는, 사용자 1이 웨어러블 스마트 디바이스(102(3)(A))를 통해 바라보는 경우에, 사용자 2의 아바타(204(2))를 보고 있는 것과, 사용자 2가 웨어러블 스마트 디바이스(102(3)(B))를 통해 바라보는 경우에, 사용자 1의 아바타(204(1))를 보는 것을 나타내고 있다. 정리하면, 웨어러블 스마트 디바이스(102(3)(A), 102(3)(B))는, 사용자 그들 자신이 보여지지 않더라도 사용자 1과 사용자 2의 시각으로부터 묘사된다. 웨어러블 스마트 디바이스(102(3)(A), 102(3)(B))는 착용자의(예컨대, 사용자의) 아바타를 제어하는데 이용되는 착용자에 관한 이미지 정보를 캡처할 수 있다. 아바타는 다른 사용자의 아바타를 제어하는 이미지 데이터를 동시에 캡처하는 그 다른 사용자의 웨어러블 스마트 디바이스에 의해 보여진다. 다르게 말하면, 사용자 2의 아바타(204(2))는 웨어러블 스마트 디바이스(102(3)(A))에 의해 사용자 1에게 표시된다. 마찬가지로, 사용자 1의 아바타(204(1))는 웨어러블 스마트 디바이스(102(3)(B))에 의해 사용자 2에게 표시된다. 이에, 본 구현예는 비디오 호출 및/또는 다른 사용을 용이하게 하는 독립형(standalone) 웨어러블 스마트 디바이스를 제공할 수 있다.

[0016] 도 4와 도 5는 사용자 3의 정면도와 측면도를 각각 보여준다. 이 경우에, '사용자 3'은 웨어러블 스마트 디바이스(102(3))(예컨대, 스마트 안경)와 웨어러블 스마트 디바이스(102(4))(예컨대, 스마트 클립온 마이크)를 착용하고 있다. 이 경우에, 사용자 3은 설명의 목적상 2개의 웨어러블 스마트 디바이스를 착용하고 있다. 대안으로, 사용자는 이들 웨어러블 스마트 디바이스 중 어느 하나를 다른 것 없이 이용할 수 있다. 도 4와 도 5는 본 설명에서의 용어 "로우 앵글 카메라"가 의미하는 것의 일례를 설명하는 것을 도와줄 수 있는 기준면(402)도 나타내고 있다. 이 구현예에서는, 기준면(402)이 사용자의 눈과 사용자의 입의 정면을 통과하거나 그 정면을 포함한다.

[0017] 이 예에 있어서, 로우 앵글 카메라(104)는 기준면(402)에 대해 약 +/- 45도 미만의 각도로 배향될 수 있다(예컨대, 로우 앵글 카메라는 사용자의 얼굴의 표면에 대체로 평행한 이미지를 캡처할 수 있다). 반면에, 스마트 폰, 태블릿, 또는 화상 회의 시스템 상에 채택되는 것과 같은 전통적인 카메라는 기준면(402)에 대체로 수직이게 배향되도록 구성되며, 로우 앵글로부터 의미있는 이미지 정보를 캡처할 수 없다.

[0018] 이 예에서, 웨어러블 스마트 디바이스(102(3))의 로우 앵글 카메라(104(3))는 대체로 기준면(402) 내에 또는 기준면에 근접하게 배치되며, 화살표(404)가 나타내는 이미지를, 기준면을 따르는 방향으로 또는 기준면에 평행하게 캡처하도록 구성된다. 마찬가지로, 웨어러블 스마트 디바이스(102(4))의 로우 앵글 카메라(104(4))는 대체로 기준면(402) 내에 또는 기준면에 근접하게 배치되며, 화살표(406)가 표시하는 이미지를, 기준면(402)을 따르는 방향으로 또는 기준면(402)에 평행하게 캡처하도록 구성된다.

[0019] 도 6은 본 개념을 달성하는 예시적인 프로세스 흐름(600)을 보여준다. 설명의 목적상, 프로세스 흐름(600)은 트레이닝 스테이지(602)와 전개 스테이지(deployment stage)(604)로 나뉘질 수 있다. 트레이닝 스테이지는 606에서 사용자에게 대한 아바타 모델을 구축할 수 있다. 전술한 바와 같이, 아바타는 사진과 동일하거나 기타 형태일 수도 있다. 사용자의 아바타는 아바타 모델(608)을 통해 형성될 수 있다. 아바타 모델은 시선, 입 움직임 등에 대응하는 것과 같은, 다수의 얼굴 파라미터(facial parameter) 또는 애니메이션 파라미터를 포함할 수 있다. 또한, 프로세스는 610에서 센서 입력과 아바타 파라미터 간의 상관에 대해 트레이닝하여 매핑(612)을 생성할 수 있다. 프로세스 흐름의 이 부분은 웨어러블 스마트 디바이스(예컨대, 웨어러블 스마트 디바이스 상의 센서)와 적색, 청색, 녹색과 깊이(RGBD) 카메라 등의 다른 촬상 디바이스로 사용자를 동시에 기록(recording)함으로써 달성될 수 있다. RGBD 카메라는 정규(도 4의 기준면(402)에 대체로 수직) 배향에 있을 수 있다. 사용자 액션이 로우 앵글 카메라, 카메라, 및/또는 다른 센서(도 1 참조) 등의 웨어러블 스마트 디바이스에 의해 그리고 RGBD 카메라에 의해 캡처될 수 있다. 매핑(612)은 RGBD 카메라와 웨어러블 스마트 디바이스에 의해 캡처되는 것인 사용자의 특정 사용자 액션을, 특정 아바타 애니메이션 파라미터에 관련시킬 수 있다.

[0020] 전개 스테이지(604)는 웨어러블 스마트 디바이스만으로 달성될 수 있다. 이 경우에, 다양한 유형의 센서 데이터(614)가 웨어러블 스마트 디바이스에 의해 수집될 수 있다. 이 예에서는, 센서 데이터(614)가 웨어러블 스마트 디바이스의 센서 구성에 따라, 무엇보다도, 비디오 데이터, 오디오 데이터, 및 접촉 데이터를 포함할 수 있다. (접촉 데이터 데이터는 피부 움직임, 근육 수축 등을 감지하는, 사용자의 피부와 접촉한 센서에 의해 수집될 수 있다.) 센서 데이터(614)는 센서 입력(618)을 도출하기 위해 616에서 프로세싱될 수 있다. 센서 입력(618)은 트레이닝 스테이지(602)로부터의 아바타 모델(608)과 매핑(612)을 이용하는 620에서 아바타 애니메이션 파라미터로 매핑될 수 있다. 이 양태는 전개 스테이지(604)에 대해 아바타 애니메이션 파라미터(622)를 출력한다. 아바

타 애니메이션 파라미터(620)(및 아바타 모델(608))은 626에서의 원격 아바타 렌더링 및 애니메이션을 위해 네트워크(626)를 통해 전송될 수 있다. 프로세스 흐름(600)의 특정 양태에 대해서는 도 7 내지 도 12를 참조하여 이하에서 상세하게 설명한다.

- [0021] 정리하면, 본 개념은 신체에 장착된, 내향(inward looking)(예컨대, 시선 추적) 및/또는 로우 앵글 카메라에 기초한 얼굴 표정 추적용 시스템을 필요로 할 수 있다. 이들 카메라로부터의 이미지는 사용자가 원격 상대방에 애니메이션화된 아바타로 표현될 수 있도록 활용될 수 있다.
- [0022] 도 7은 웨어러블 스마트 디바이스(102(3))로부터의 전체 얼굴 표정 추적을 제공할 수 있는 예시적인 시스템(700)을 나타내고 있다. (도 7은 도 6과 다소 중복되지만, 도 6에 대해 도입된 개념의 설명을 돕는 상이한 강조점을 갖고 있다.) 설명의 목적상, 시스템(700)은 입력열(701(1))과 출력열(701(2))로 나누어진다.
- [0023] 이 경우에, 시스템(700)은 가시광 카메라(예컨대, 흑색과 백색 또는 적색, 녹색, 청색(RGB)) 또는 적색, 녹색, 청색 + 깊이(RGBD) 카메라(702) 등의 기준 카메라를 포함한다. 이러한 깊이 카메라의 일례는 Microsoft®사에서 공급하는 Kinect® 브랜드의 깊이 카메라이다. RGBD 카메라(702)는 전형적으로 사용자쪽에 배향될(예컨대, 도 4와 도 5에 대해 전술한 기준면에 수직인 사용자에게 대체로 지향될) 수 있다. (도 7의 예시에서는, 사용자가 정면을 향해 있고, RGBD 카메라(702)가 사용자의 측면에 있음을 주목해야 한다. RGBD 카메라는 사실상 사용자의 정면에 있게 될 것이다(예컨대, 사용자의 전체 얼굴 이미지를 캡처하도록 배치).)
- [0024] 트레이닝 스테이지(602) 중에, RGBD(702)는 전체 얼굴 이미지(704)를 캡처할 수 있다. 마찬가지로, 웨어러블 스마트 디바이스의 로우 앵글 카메라(104(3))와 시선 추적 카메라(106(3))는 RGBD 카메라(702)와 동시에 사용자의 각각의 이미지(706, 708)를 캡처한다. 도면의 스케일 때문에, 이들 엘리먼트가 도 7에서는 쉽게 구별될 수 없지만, 도 1, 도 4 및 도 5에서는 수월하게 알아볼 수 있다. 로우 앵글 카메라(104(3)) 및 시선 추적 카메라(106(3))의 제한된 시야로 인해, 사용자의 얼굴의 일부만 각 카메라(104(3), 106(3))에 보여진다. 구체적으로, 로우 앵글 카메라(104(3))는 이미지(706)가 나타내는 바와 같이 입과 턱쪽으로 뺨 아래에 지향되고, 시선 추적 카메라(106(3))는 이미지(708)가 나타내는 바와 같이 사용자의 눈쪽으로 지향된다. (예시하는 이미지는 설명의 목적상 표시되는 것이며 제한적인 것으로 의도되지 않음을 알아야 한다.) 또한, 로우 앵글 카메라에 채택되는 렌즈의 유형은 이미지에 그리고 사용자에게 대한 배향 각도에도 영향을 미칠 수 있다. 이를테면, 어안 렌즈는 표준 렌즈보다도 소정의 방향으로 사용자의 얼굴을 더 많이 캡처할 수 있다. 이러한 구현예에서는, 이미지(706)에서 눈과 입을 캡처하지만 이미지의 2개 세트를 사용하지 않는(예컨대, 이들을 캡처하는 카메라 또는 이미지(708)를 사용하지 않는) 단일 로우 앵글 카메라(104(3))가 사용될 수 있다.
- [0025] 또 다른 구현예에서는, 다른 센서 데이터가 로우 앵글 카메라(104(3))로부터의 이미지 데이터를 증강시킬 수 있다. 이를테면, 도 6에서는 오디오 데이터 및/또는 사용자 접촉 데이터 등의, 이미지(예컨대, 비디오) 외에 다양한 다른 형태의 센서 데이터(614)를 나타내었음을 상기해야 한다. 이러한 경우에, 로우 앵글 카메라(104(3))는 스마트 웨어러블 디바이스 상의 유일한 카메라일 수 있다. 로우 앵글 카메라(104(3))는 예컨대 사용자의 입만 아니라 사용자의 눈과 뺨을 캡처할 수도 있다. 이러한 경우에, 마이크(도 1의 마이크(108) 참조)로부터의 오디오 데이터는 모델 구축(606)에 사용될 수 있는 오디오 데이터를 캡처할 수 있다. 이를테면, 사용자가 말하는 개별 사운드는 트레이닝 스테이지(602) 중에 입 움직임 및/또는 얼굴 표면에 관한 개별 아바타 파라미터 값으로 매핑될 수 있다. 이에, 오디오 데이터는 아바타 파라미터의 일부에 사용될 수 있고, 로우 앵글 카메라(104(3))로부터의 이미지 데이터는 다른 아바타 파라미터에 이용될 수 있다.
- [0026] 예시하는 구성에서는, 트레이닝 스테이지(602) 중에, 웨어러블 스마트 디바이스(102(3))의 2개의 카메라(104(3), 106(3))로부터의 부분적 얼굴 이미지 스트림이, 전체 얼굴 이미지(704)의 견지에서 평가될 때에 사용자의 전체 얼굴 표정을 추정하는데 이용될 수 있다. 이 추정은 다수의 얼굴 표정 파라미터에 대한 정보를 포함할 수 있다. 다수의 얼굴 표정 파라미터는 사용자의 애니메이션화 아바타(710)를 구동하는데 이용될 수 있다. 이를테면, 일부 구현예에서는 깊이에 기초한 얼굴 표정 추적 알고리즘을 채택하여 사용자에게 대한 트레이닝 데이터를 자동으로 수집할 수 있다. 그리고 나서, 트레이닝 데이터로부터 얼굴 표정 파라미터를 추정하기 위해 다수의 머신 러닝 알고리즘(machine learning algorithm)을 모델을 구축하는데 채택할 수 있다. 다수의 가능한 기술들 가운데, 무엇보다도, 리지 회귀(ridge regression), 멀티태스크 피쳐 러닝(multitask feature learning), 및/또는 통상의 신경망이 채택될 수 있다.
- [0027] 다르게 말하면, 트레이닝 스테이지(602) 중에, 웨어러블 스마트 디바이스(102(3))에 의해 캡처된 트레이닝 이미지 세트(예컨대, 비디오) 및 대응하는 RGBD 카메라(702)로부터의 실측(ground truth) 얼굴 표정(예컨대, 이미지(704))이 이들의 관계를 기술하는 아바타 모델(도 6의 608)을 구축하는데 이용될 수 있다. 본 구현예에서는 실

측 이미지가 RGBD 카메라로부터 취득되지만, 다른 구현예에서는 다른 유형의 센서 또는 카메라를 이용할 수 있음을 알아야 한다. 이를테면, 실측 이미지는 표준 RGB 카메라, 적외선 카메라, 및/또는 다른 유형의 센서에 의해 취득될 수 있다. 전개 스테이지(604)에서는, 웨어러블 스마트 디바이스로부터의 입력 이미지(706(1), 708(1))(예컨대, 비디오 페어(pair) 스트림)를 얼굴 표정 파라미터의 세트로 매핑하는데 모델이 이용되는데, 이 얼굴 표정 파라미터의 세트는 원격 상대방에서 아바타(710(1))를 구동하는데 이용될 수 있다. 도 6와 도 7에 대해 앞에서 소개한 특정 양태들에 대해 이하에 상세하게 설명한다.

[0028] **데이터 취득**

[0029] 도 4와 도 5로부터, 본 구현예에서는, 웨어러블 스마트 디바이스(102(3))의 카메라가 비대칭으로 배치되는 것을 상기해야 한다(예컨대, 로우 앵글 카메라(104(3))는 입쪽으로 얼굴에 평행하게 지향되고, 시선 추적 카메라(106(3))는 사용자의 눈에 지향된다). 비대칭성에 대한 유도(motivation)는, 중복되는 얼굴 부분을 캡처하는 데에는 하나의 카메라만이 채택되고, 뺨 등의 비대칭일 수 있는 얼굴의 부분들은 양쪽 카메라에 의해 캡처된다는 것이다. 물론, 대칭 및/또는 비대칭 구성으로 된 3개 이상의 카메라 등의 다른 구성도 고려된다.

[0030] 도 7을 다시 참조하면, RGBD 카메라 입력(예컨대, 이미지(704))으로부터 얼굴 표정을 추적하는 데에 다양한 접근방법을 이용할 수 있다. 이를테면, 무엇보다도 Microsoft사의 Kinect® 아바타 추적 알고리즘 등의 RGBD 추적 알고리즘이 이용될 수 있다. 일부 RGBD 추적 알고리즘 구현에는 개별화된 메시 모델의 세트 $B = \{B_0, \dots, B_n\}$ 를 수반한다는 가정 아래 유효할 수 있으며, 여기서 B_0 는 정지 자세(rest pose)이고, $B_i (i > 0)$ 는 정지 자세와 어떤 얼굴 표정 사이의 차이를 표현하는 추가 변위량이다. 이에 임의의 얼굴 표정은 다음과 같이 생성될 수 있다.

$$M_j = B_0 + \sum_{i=1}^n y_{ij} B_i, \quad (1)$$

[0032] 여기서, y_{ij} 는 자세 M_j 에 대한 혼합 가중치(blending weight)이다.

[0033] RGBD 추적 알고리즘은 48개의 혼합 형상(blend shape) 등의 다수의 혼합 형상을 이용할 수 있고, 이에 캡처된 프레임마다 48개의 혼합 가중치를 생성할 것이다. 따라서, 얼굴 표정을 추정하는 태스크는 48개의 회귀 함수를 학습하는 문제로 재검토될 수 있으며, 그 함수 각각은 웨어러블 스마트 디바이스 이미지로부터 수신된 정보를 대응하는 혼합 형상 가중치에 매핑시킨다.

[0034] 이어지는 설명은 이미지 사전 프로세싱에 의해 개시되는 얼굴 표정 추적에 관한 것이다.

[0035] 일부 구현예는, 사용자가 입는 서로 다른 옷들, 웨어러블 스마트 디바이스를 착용하고 벗을 때의 작은 움직임, 및 사용자들 간의 차이에 대한 알고리즘의 로버스트성(robustness)을 증가시키기 위해 입력 이미지 데이터에 대해 사전 프로세싱을 수행할 수 있다.

[0036] 제1 유형의 사전 프로세싱은 피부 필터링에 관련될 수 있다. 이 프로세스는 3개의 단계에 기초할 수 있다. 첫째, 피부가 되는 각 픽셀의 확률을 추정하기 위해 간단한 분류기(classifier)가 이용될 수 있다. 픽셀 분류만 이용하는 것은, 일부 긍정 오류(false positive) 픽셀이 손상될 수 있을 때에, 특히 사지가 배경에 나올 때에 충분히 로버스트적일 수 없다. 프로세스를 더욱 로버스트적이게 하기 위해, 일부 구현예에서는 이미지들의 엣지를 검출하고, 이들 엣지 중에서, 피부와 배경 영역을 분리시키는 (가능하다면) 최적의 것을 찾을 수 있다.

[0037] 그렇게 하기 위해, 이들 구현예는 전체 이미지에 대해 먼저 캐니 엣지(Canny edge) (또는 다른) 검출기를 적용할 수 있다. 검출기의 감도(sensitivity)는 다수의 엣지 세트가 검출되고 임의의 유용한 엣지가 그 세트에 포함되도록 높게 설정될 수 있다. 마지막으로, 이들 구현예에서는 이전의 2개의 프로세스에서 수집된 정보를 활용하는 다이내믹 프로그래밍 알고리즘을 이용하여 (가능하다면) 최적의 엣지를 찾을 수 있다. 이 방법은 예컨대 좌측에서부터 우측으로 이미지 열(column)에 대해 반복될 수 있다. 각 열마다, 프로세스는 엣지에 속하는 각 픽셀에 대해, 그 픽셀이 (가능하다면) 최적의 엣지의 부분일 확률을 산출할 수 있다. 이것은 그 픽셀 위 아래의 픽셀들이 피부일 확률과, 이전의 반복에서 산출되었던 좌측에 이웃하는 픽셀의 가능성에 기초하여 산출될 수 있다.

[0038] 도 8은 이러한 일례를 보여준다. 실례 1은 프로세싱 전의 내향 시선 추적 카메라(106(3))(도 4와 도 5)로부터의 이미지(800(1))를 보여준다. 이 설명은 라인 드로잉(line drawing)으로 묘사하기에 어려운 색상을 다루어야 한다. 실례 1에 있어서, 교차빔금(cross-hatching)으로 표시된 영역(802)은 인접한 피부(804)와 같은 색상에 속하

는 배경이며, 이 경우에도 구분하기가 곤란할 수 있다. 경계(806)가 검출될 수 있고, 실례 2에 있어서 배경 영역의 색상은 프로세싱된 이미지(800(2)) 내의 808에 나타내는 바와 같이 대비색(이 경우에는 대각선으로 표시되는 흰색)으로 변한다.

[0039] 도 9는 일부 구현예에서 채택되는 사전 프로세싱의 제2 유형의 일례를 보여준다. 이 제2 유형의 사전 프로세싱은 카메라의 작은 움직임 또는 상이한 사용자들 간의 차이를 처리할 수 있다. 이러한 경우에, 템플릿 매칭에 기초하여 모든 이미지(이 경우에, 좌측 이미지(900(1))와 우측 이미지(900(2)))를 얼라인(aline)하는 레지스트레이션 방법(registration method)이 채택될 수 있다. 코-팸 주름(902)과 외측 오른쪽 눈 구석(904)이 좌측 및 우측 카메라 이미지(900(1), 900(2))에서 각각 탐색되고, 이미지들 간의 레지스트레이션 에러를 줄이기 위해 입력 이미지가 그에 따라 조정된다. 실험 결과에 나타내는 바와 같이, 이 레지스트레이션 단계는 새로운 사용자를 트레이닝하는데 매우 유용할 수 있다.

[0040] 사전 프로세싱 후에, 입력 이미지의 2개의 형태, 즉 로 픽셀(raw pixel) 및 로컬 바이너스 패턴(local binary pattern, LBP) 묘사(description)가 이용될 수 있다.

[0041] 설명의 목적상, 트레이닝 입력 데이터를, 행렬 $X \in \mathbb{R}^{d \times N}$ (차원수 d 를 가진 N 개의 트레이닝 인스턴스를 포함)와 $Y \in \mathbb{R}^{N \times 48}$ (트레이닝 레벨(RGBD 추적 알고리즘으로부터 취득된 혼합 형상 가중치))로 표시한다. y_t 는 Y 의 t 번째 열, 즉 t 번째 혼합 형상에 대한 실측 가중치(ground truth weight)를 표시하는데 이용될 수 있다. 머신러닝 문제가 다중 출력 회귀에 존재할 수 있다면, 3개의 머신러닝 알고리즘(리지 회귀, 멀티태스크 피쳐 러닝, 및 컨볼루션 신경망(convolutional neural network)이 채택되고 있다는 것이다. 물론, 다른 머신러닝 알고리즘도 다른 구현예에 채택될 수 있다.

[0042] 리지 회귀

[0043] 리지 회귀(Ridge Regression, RR)는 다음의 최적화 문제를 해결함으로써 출력 각각마다 선형 모델을 학습할 수 있다.

$$\underset{w_t}{\operatorname{argmin}} \|X^T w_t - y_t\|_2^2 + \lambda \|w_t\|_2^2, \quad (2)$$

[0045] 여기서, w_t 는 혼합 형상 t 에 대한 d 차원의 가중 벡터이다. 전술한 문제는 매 혼합 형상마다 독립적으로 해결된다.

[0046] 멀티태스크 피쳐 러닝

[0047] 혼합 형상 가중치들이 얼굴 표정을 묘사하기에 조직적이고, 이들 혼합 형상이 비직교적이라면, 이들은 선형으로 상관될 것으로 예측될 수 있다. 그 결과, 멀티태스크 피쳐 러닝(multi-task feature learning, MTL)이 본 태스크에 채택될 수 있다. MTL은 모든 태스크를 동시를 학습함으로써 혼합 형상 가중치들 사이에 공통성을 이용하려고 할 수 있다. MTL은 다음의 최적화 문제를 해결하는 것을 시도할 수 있다.

$$\underset{C, D}{\operatorname{argmin}} \|X^T DC - Y\|_{Fro}^2 + \frac{\lambda}{2} (\|C\|_{Fro}^2 + \|D\|_{Fro}^2), \quad (3)$$

[0049] 여기서, $\|\cdot\|_{Fro}^2$ 은 행렬의 프로베니우스노름(Frobenius norm)의 제곱, 즉 행렬의 요소들의 제곱의 합을 표시한다. 도 10에 도시하는 바와 같이, 입력을, 공유 피쳐(shared feature)(1002)를 나타내는 행렬 D 와 곱하여 형성된 모든 태스크에 대해 공통 레이어(common layer)가 있을 수 있다. 그래서, 출력 레이어(1004)는 공유 피쳐를 행렬 C 와 곱해서 형성된다. 양쪽 레이어는 선형이다.

[0050] 수식 (3)의 문제는 비볼록(non-convex)이다. 대신에 프로세스는 이하에 주어진 그것의 등가의 볼록한 문제를 최적화한다.

$$\underset{W}{\operatorname{argmin}} \quad \|X^T W - Y\|_{Fro}^2 + \lambda \|W\|_{Tr}, \quad (4)$$

여기서, $W = DC$ 이고, $\|\cdot\|_{Tr}$ 는 행렬의 트레이스노름(trace norm), 즉 그것의 특이값(singular value)의 합을 표시한다.

신경망

도 11a와 도 11b는 2개의 컨볼루션 신경망(1100(1), 1100(2))을 보여준다. 제1 컨볼루션 신경망(1100(1))은 컨볼루션 레이어(convolutional layer)(110(1))에서 입력 이미지(102(1))를 수신한다. 이 예에서는, 컨볼루션 레이어(110(1))가 사이즈 3×3 의 16개의 커널(kernel)을 갖는다. 컨볼루션 레이어의 출력은 드롭아웃(dropout, DO) 가능하며, 50 노드의 레이어(1106(1))에 완전 연결된다. 50 노드 레이어로부터의 출력은 이어서 최종 출력(1110)을 생성하는 48 노드의 출력 레이어(1108(1))에 전송된다.

제2 컨볼루션 신경망(1100(2))은, 역시 사이즈 3×3 의 16개의 커널을 가진 컨볼루션 레이어(1104(2))인 제1 레이어를 포함한다. 제1 컨볼루션 레이어(1104(2))는 입력 이미지(1102(2))를 수신한다. 제1 컨볼루션 레이어의 출력은 제2 은닉 레이어(1106(2))에 밀집 연결된다. 이 예에서는 제2 은닉 레이어가 200 은닉 유닛을 가진다. 제2 은닉 레이어의 출력은 제3 레이어(1108(2))로 향한다. 이 예에서는, 제3 레이어가 완전 연결 레이어로서, 또한 200 출력 노드를 갖는다. 이 경우에, 최종 레이어(1112)는 제3 레이어(1108(2))의 출력을 수신하여 최종 출력(1114)을 생성한다. 제3 레이어(1108(2))는 48개의 출력을 가진 완전 연결 레이어일 수 있다. 일부 구현예에서는, ReLU(rectified linear unit) 비선형성이, 출력 레이어를 제외한 모든 컨볼루션 및 완전 연결 레이어에 이용된다. 또한, 완전 연결 레이어의 일부에서 드롭아웃이 가능하다.

정리하면, 이상에서는, 사용자의 얼굴에 거의 평행한 경우와 같이, 일반적이지 않은 배향으로 배치되는 웨어러블 카메라(예컨대, 웨어러블 스마트 디바이스의 카메라)로부터 얼굴 표정 추적을 수행하는 방법을 설명하였다. 실측 데이터는 웨어러블 스마트 디바이스의 카메라와 동시에 깊이 카메라를 이용하여 취득될 수 있다. 깊이 기초한 얼굴 표정 추정 알고리즘은 실측 데이터에서 2개의 카메라로부터의 이미지 정보를 상관시키는데 이용될 수 있다. 실측 데이터는 얼굴 파라미터와 연관된 회귀 모델을 트레이닝하는데 이용될 수 있다. 그 후에, 사용자의 아바타를 제어하는 파라미터를 도출하기 위해 웨어러블 스마트 디바이스로부터의 이미지가 회귀 모델로 프로세싱될 수 있다.

도 12는 시각화 시스템(1200)을 보여준다. 설명의 목적상, 시스템(1200)은 스마트 안경임이 자명한 2개의 웨어러블 스마트 디바이스(102(3)(A), 102(3)(B))를 포함한다. 이들 예시하는 웨어러블 스마트 디바이스는 2개의 디바이스 구성을 대표하는데, 이 구성 중 어느 하나는 스마트 안경 이외의 다른 웨어러블 스마트 디바이스에 적용될 수도 있다. 간단하게는, 웨어러블 스마트 디바이스(102(3)(A))는 운영체제 중심의 구성을 대표하고, 웨어러블 스마트 디바이스(102(3)(B))는 시스템온칩(system on chip) 구성을 대표한다. 이들 양태에 대해서는 이하에서 보다 상세하게 설명할 것이다. 시스템(1200)은 웨어러블 스마트 디바이스가 아닌 디바이스(1202)도 포함한다. 이 경우에, 디바이스(1202)는 스마트 폰이 자명하나, 웨어러블 스마트 디바이스와 통신할 수 있는 임의의 디바이스를 대표하는 것이 의도된다. 시스템(1200)은, 네트워크(1206)를 통해 웨어러블 스마트 디바이스와 통신할 수 있는 클라우드 기반의 리소스 등의 원격 리소스(1204)도 포함한다. 디바이스(1202)와 원격 리소스(1204) 중 어느 하나 또는 양쪽은 본 개념을 달성하기 위하여 웨어러블 스마트 디바이스와 협력으로 동작할 수 있다. 웨어러블 디바이스(102(3)(A), 102(3)(B)), 디바이스(1202), 및/또는 원격 리소스(1204) 중 임의의 것은 아바타 제어 컴포넌트(1208)를 지원하거나 포함할 수 있다. 아바타 제어 컴포넌트는 웨어러블 스마트 디바이스(102(3)(A), 102(3)(B))의 다른 엘리먼트의 설명 후에 이하에서 보다 상세하게 설명할 것이다.

설명 목적상, 웨어러블 스마트 디바이스(102(3)(A))는 애플리케이션(1210), 운영체제(1212) 및 하드웨어(1214)로 구성될 수 있다. 웨어러블 스마트 디바이스(102(3)(B))는 공유 리소스(1216), 전용 리소스(1218), 및 그 사이의 인터페이스(1220)로 구성될 수 있다.

웨어러블 스마트 디바이스(102(3)(A), 102(3)(B))는, 프레임(1222)을 포함할 수 있고 템플릿(1224)에 접속되는 배향 구조(100)를 더 포함할 수 있다. 템플릿은 장축(x 기준축에 평행)을 따라 연장되어 귀부분(earpiece)(1226)에서 끝날 수 있다. 프레임은 (한 쌍의) 렌즈(1228)를 지지할 수 있다. 웨어러블 스마트 디바이스는 스토리지(1230), 프로세서(1232), 배터리(1234)(또는 다른 전원), 센서(1236), 및/또는 통신 컴포넌트(1238)를 더 포함할 수 있다. 웨어러블 스마트 디바이스는 간결함을 위해 여기에 도시하거나 설명하지 않는, 임

출력 디바이스, 버스, 그래픽 카드(예컨대, 그래픽 프로세싱 유닛(graphics processing unit, GPU) 등의 기타 엘리먼트를 대안적으로 또는 부가적으로 포함할 수 있다. 일 구성에 있어서, 배터리(1234)는 쿼부분(1226) 중 한쪽에 배치될 수 있고, 통신 컴포넌트(1238), 스토리지(1230), 및 프로세서(1232)는 다른쪽 안경 다리에 배치된다.

[0060] 본 예에 있어서, 센서(1236)는 로우 앵글 카메라(104), 시선 추적 카메라(106), 및 마이크(108)임이 자명하다. 본 명세서에서 사용될 때에, 카메라는 사용자의 얼굴에 대한 정보를 캡처할 수 있는 임의 유형의 센서로서 간주될 수 있다. 카메라는 가시광 또는 기타 파장일 수 있다. 카메라는 상이한 유형의 렌즈를 채택할 수 있다. 이를테면, 카메라는 무엇보다도 볼록 렌즈, 비구면 렌즈, 및/또는 어안 렌즈를 채택할 수 있다. 어안 렌즈는 단일 카메라로 하여금 사용자의 얼굴의 큰 부분을 캡처하게 하는데, 다른 렌즈의 경우에는 이 부분을 캡처하기 위해서 2개의 카메라를 필요로 할 수 있다. 카메라에 대한 추가 상세내용은 앞에서 도 1의 도입부에 설명하였다. 물론, 카메라가 상세하게 설명되지만, 다른 구현에도 대안으로 또는 추가적으로 다른 유형의 센서(1236)를 사용할 수 있다.

[0061] 안경 렌즈(1228)는 교정 렌즈이거나 비교정 렌즈일 수 있다. 일부 구현예에 있어서, 렌즈는 사용자가 바라보는 이미지를 표시할 수 있는 기능을 가질 수 있다. 이러한 표시 기능은 투사에 의해, 또는 렌즈 내에 발광 다이오드나 다른 컴포넌트를 포함함으로써 달성될 수 있다.

[0062] 통신 컴포넌트(1238)는 셀룰러, Wi-Fi(IEEE 802.xx), 블루투스 등의 다양한 기술로 통신하는, 수신기와 송신기 및/또는 기타 무선 주파수 회로를 포함할 수 있다.

[0063] 아바타 제어 컴포넌트(1208)는 센서(1236)로부터 입력을 수신할 수 있다. 아바타 제어 컴포넌트(1208)는 센서 데이터로부터 애니메이션 파라미터를 도출하여 사용자의 아바타를 제어하기 위해 아바타 모델을 활용할 수 있다. 개별 웨어러블 스마트 디바이스(102(3)) 상의 아바타 제어 컴포넌트(1208)는 애니메이션 파라미터의 값을 결정하기 위해 센서 데이터를 프로세싱할 정도로 비교적 로버스트적일 수 있다. 예를 들어, 웨어러블 스마트 디바이스(102(3)(A))는 센서(1236(1))로부터 애니메이션 파라미터 값을 결정할 수 있다. 웨어러블 스마트 디바이스(102(3)(A))는 사용자의 아바타를 애니메이션화하여 그 애니메이션을 다른 디바이스에 전송할 수 있거나, 실행을 위해 애니메이션 파라미터 값을 다른 디바이스에 전송할 수 있다.

[0064] 이를테면, 웨어러블 스마트 디바이스(102(3)(A))에 있어서, 아바타 제어 컴포넌트(1208(1))는 웨어러블 스마트 디바이스(102(3)(B)) 등의 원격 디바이스와의 통신을 위해 애니메이션 파라미터 값을 통신 컴포넌트(1238(1))에 전송할 수 있다. 수신측 디바이스의 아바타 제어 컴포넌트(1208(2))는 수신된 애니메이션 파라미터 값에 기초하여 사용자의 아바타를 제어할 수 있다. 동시에, 원격 디바이스의 아바타 제어 컴포넌트(1208(2))는 센서(1236(2))로부터 데이터를 수신하고, 애니메이션 파라미터 값을 결정하기 위해 센서 데이터를 프로세싱하며, 또 애니메이션 파라미터 값을 웨어러블 스마트 디바이스(102(3)(A))에 전송할 수 있다.

[0065] 다른 구성에 있어서, 웨어러블 스마트 디바이스(102(3)(A))는 덜 로버스트적일 수 있고, 아바타 제어 컴포넌트(1208(3) 또는 1208(4)) 각각에 의한 프로세싱을 위해 센서 데이터의 일부를 디바이스(1202) 또는 원격 리소스(1204)에 전송할 수 있다. 이들 아바타 제어 컴포넌트(1208(3) 또는 1208(4))는 웨어러블 스마트 디바이스(102(3)(A))를 대신하여 애니메이션 파라미터 값을 결정할 수 있고, 웨어러블 스마트 디바이스(102(3)(A))를 대신하여 이들 값을 웨어러블 스마트 디바이스(102(3)(B))에 전송할 수 있다.

[0066] 또 다른 구성에서는 웨어러블 스마트 디바이스(102(3)(A))가 애니메이션 파라미터 값을 디바이스(1202)에 전송할 수도 있다. 디바이스(1202)는 (예컨대, 비디오 호출에서의) 사용자의 '통상의' 비디오 스트림을 캡처하여 그 비디오 스트림을 웨어러블 스마트 디바이스(102(3)(A))에 전송할 수도 있다. 이에, 웨어러블 스마트 디바이스(102(3)(A))의 사용자는 디바이스(1202)의 사용자의 통상의 비디오를 보게 되고, 디바이스(1202)의 사용자는 웨어러블 스마트 디바이스(102(3)(A))의 사용자의 애니메이션화된 아바타를 보게 된다.

[0067] 한 관점에서는, 웨어러블 스마트 디바이스(102(3)(A), 102(3)(B)), 디바이스(1202), 및/또는 원격 리소스(1204)가 컴퓨터로서 간주될 수 있다.

[0068] 본 명세서에서 사용하는 용어 "디바이스", "컴퓨터", 또는 "컴퓨팅 디바이스"는 어느 정도의 프로세싱 능력 및/또는 저장 능력을 갖는 임의 유형의 디바이스를 의미할 수 있다. 프로세싱 능력은 기능을 제공하기 위해 컴퓨터 관독 가능한 명령어의 형태로 데이터를 실행할 수 있는 하나 이상의 프로세서에 의해 제공될 수 있다. 컴퓨터 관독 가능한 명령어 및/또는 사용자 관련 데이터 등의 데이터가 컴퓨터의 내부 또는 외부에 있을 수 있는, 스토리지/메모리 등의 스토리지 상에 저장될 수 있다. 스토리지는 무엇보다도 휘발성 또는 비휘발성 메모리, 하드

드라이브, 플래시 저장 장치, 및/또는 광학적 저장 장치(예컨대, CD, DVD 등), 원격 스토리지(예컨대, 클라우드 기반의 스토리지) 중 임의의 하나 이상을 포함할 수 있다. 본 명세서에서 사용될 때에, 용어 "컴퓨터 판독 가능한 매체"는 신호를 포함할 수 있다. 반면, 용어 "컴퓨터 판독 가능한 저장 매체"는 신호를 배제한다. 컴퓨터 판독 가능한 저장 매체는 "컴퓨터 판독 가능한 저장 디바이스"를 포함한다. 컴퓨터 판독 가능한 저장 디바이스의 예는 무엇보다도, RAM 등의 휘발성 저장 매체와, 하드 드라이브, 광디스크, 및 플래시 메모리 등의 비휘발성 저장 매체를 포함한다.

[0069] 디바이스의 예는, 퍼스널 컴퓨터, 데스크탑 컴퓨터, 노트북 컴퓨터, 셀폰, 스마트폰, 퍼스널 디지털 어시스턴트, 패드형 컴퓨터, 디지털 화이트보드, 카메라, 스마트 안경 등의 웨어러블 디바이스, 또는 계속 발전 중인 또는 이제 개발될 유형의 무수한 컴퓨팅 디바이스 중 임의의 것 등의 전통적 컴퓨팅 디바이스를 포함할 수 있다.

[0070] 전술한 바와 같이, 웨어러블 스마트 디바이스(102(3)(B))는 시스템온칩(SOC) 타입 설계로서 간주될 수 있다. 이러한 구성에 있어서, 디바이스에 의해 제공되는 기능은 단일 SOC 또는 다중 결합형 SOC 상에 통합될 수 있다. 하나 이상의 프로세서에 메모리, 스토리지 등의 공유 리소스, 및/또는 소정의 특정 기능을 수행하도록 구성된 하드웨어 블록 등의 하나 이상의 전용 리소스를 조정하도록 구성될 수 있다. 이에, 본 명세서에서 사용될 때에 용어 "프로세서"는 중앙 처리 장치(CPU), 그래픽 처리 장치(GPU), 컨트롤러, 마이크로컨트롤러, 프로세서 코어, 또는 다른 유형의 프로세싱 디바이스도 가리킬 수 있다.

[0071] 일반적으로, 본 명세서에서 설명한 기능의 일부는 소프트웨어, 펌웨어, 하드웨어(예컨대, 고정된 논리 회로), 수동 프로세싱, 또는 이들 구현의 조합을 이용해서 구현될 수 있다. 본 명세서에서 사용될 때에 용어 "컴포넌트"는 일반적으로 소프트웨어, 펌웨어, 하드웨어, 디바이스 또는 네트워크 전체, 또는 이들의 조합을 대표한다. 이를테면, 소프트웨어 구현예의 경우에, 이들은 프로세서(예컨대, CPU 또는 CPU들) 상에서 실행될 때에 특정 태스크를 수행하는 프로그램 코드를 대표할 수 있다. 프로그램 코드는 컴퓨터 판독 가능한 매체 등의 하나 이상의 컴퓨터 판독 가능한 메모리 디바이스에 저장될 수 있다. 이들 컴포넌트의 특징 및 기술은 플랫폼 독립적(platform-independent)이며, 이것은 다양한 프로세싱 구성을 구비하는 다양한 상업용 컴퓨팅 플랫폼 상에서 구현될 수 있음을 의미한다.

[0072] 본 구현예들 중 어떤 것에 있어서, 사용자가 동의를 표할 때에만 시각화 기능을 가능하게 함으로써 사용자의 프라이버시가 보호될 수 있다. 모든 프라이버시와 보안 절차가 사용자를 보호하기 위해 구현될 수 있다. 이를테면, 사용자는 권한위임(authorization)을 제공할 수 있다(그리고/또는 권한위임의 조건을 정할 수 있다). 웨어러블 스마트 디바이스는 권한위임의 조건에 따라서만 사용자의 얼굴 추적을 진행할 수 있다. 그렇지 않으면, 사용자 정보는 수집되지 않는다. 마찬가지로, 사용자는 시각화 데이터를 포함하는 자신의 데이터의 이용을 결정하는 것이 허용될 수 있다. 시각화 데이터의 어떤 사용도 정해진 사용자 조건에 부합되어야 한다.

[0073] **결론**

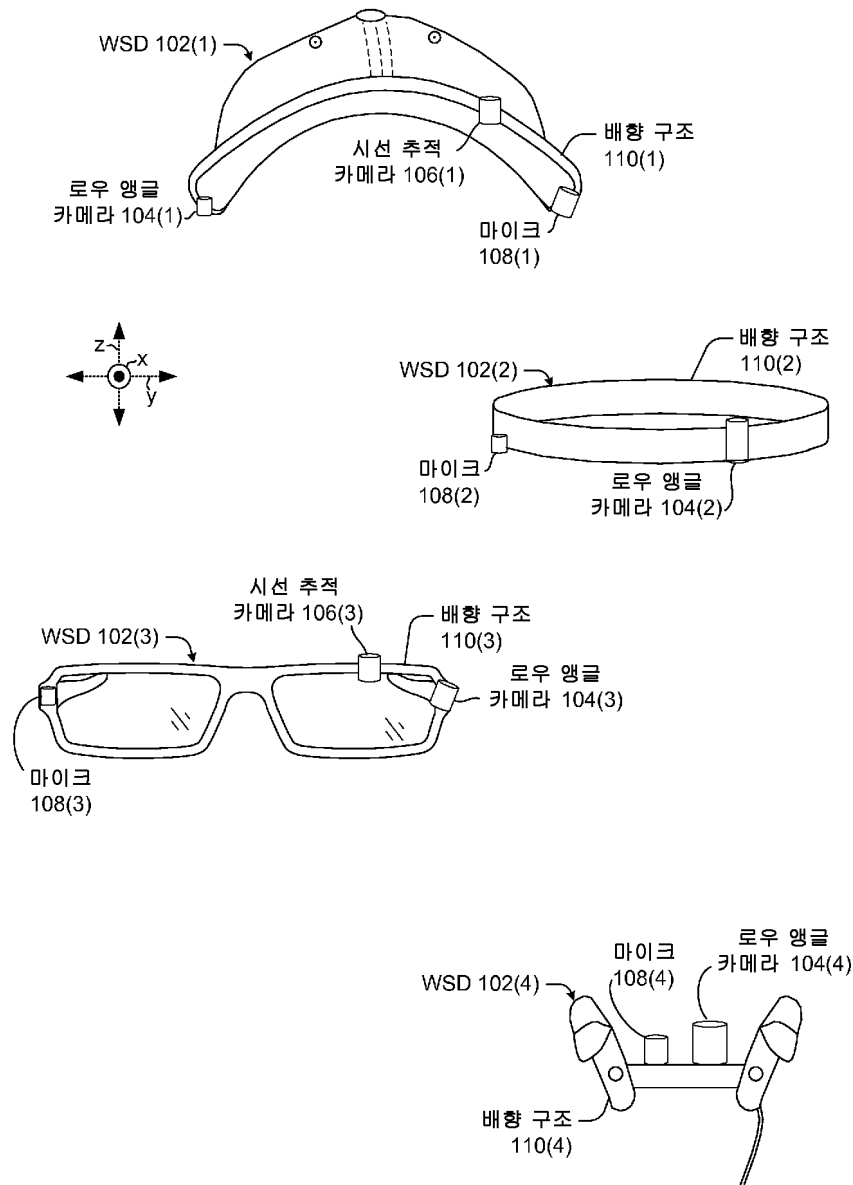
[0074] 본 구현예들은 사용자의 얼굴의 부분적 뷰만 캡처하는 로우 앵글 카메라(및/또는 기타 센서)로부터 유용한 정보를 도출할 수 있다. 본 구현예들은 Microsoft®사에서 제공하는 Kinect® 브랜드의 깊이 센서 등의 깊이 센서의 도움으로 수집된 트레이밍 데이터에 기초하여 얼굴 표정 파라미터를 추정하기 위해 머신 러닝 알고리즘을 채택할 수 있다. 본 구현예들은 카메라의 매우 제한된 뷰 각도에서도 믿을 수 있게 사람들의 얼굴 표정을 추적할 수 있다.

[0075] 설명한 방법 또는 프로세스는 전술한 시스템 및/또는 디바이스에 의해, 및/또는 기타 디바이스 및/또는 시스템에 의해 수행될 수 있다. 그 방법을 기술하는 순서는 제한으로 해석되는 것을 의도하지 않으며, 기술하는 액트의 임의 개가 그 방법 또는 대체 방법을 구현하기 위해 임의의 순서로 조합될 수도 있다. 또한, 이 방법은 디바이스가 그 방법을 구현할 수 있도록 임의의 적절한 하드웨어, 소프트웨어, 펌웨어, 또는 이들의 조합으로 구현될 수 있다. 한 경우에 있어서, 방법은 컴퓨팅 디바이스의 프로세서에 의한 실행이 컴퓨팅 디바이스가 그 방법을 수행하게 하는 명령어 세트로서 컴퓨터 판독 가능한 저장 매체 상에 저장된다.

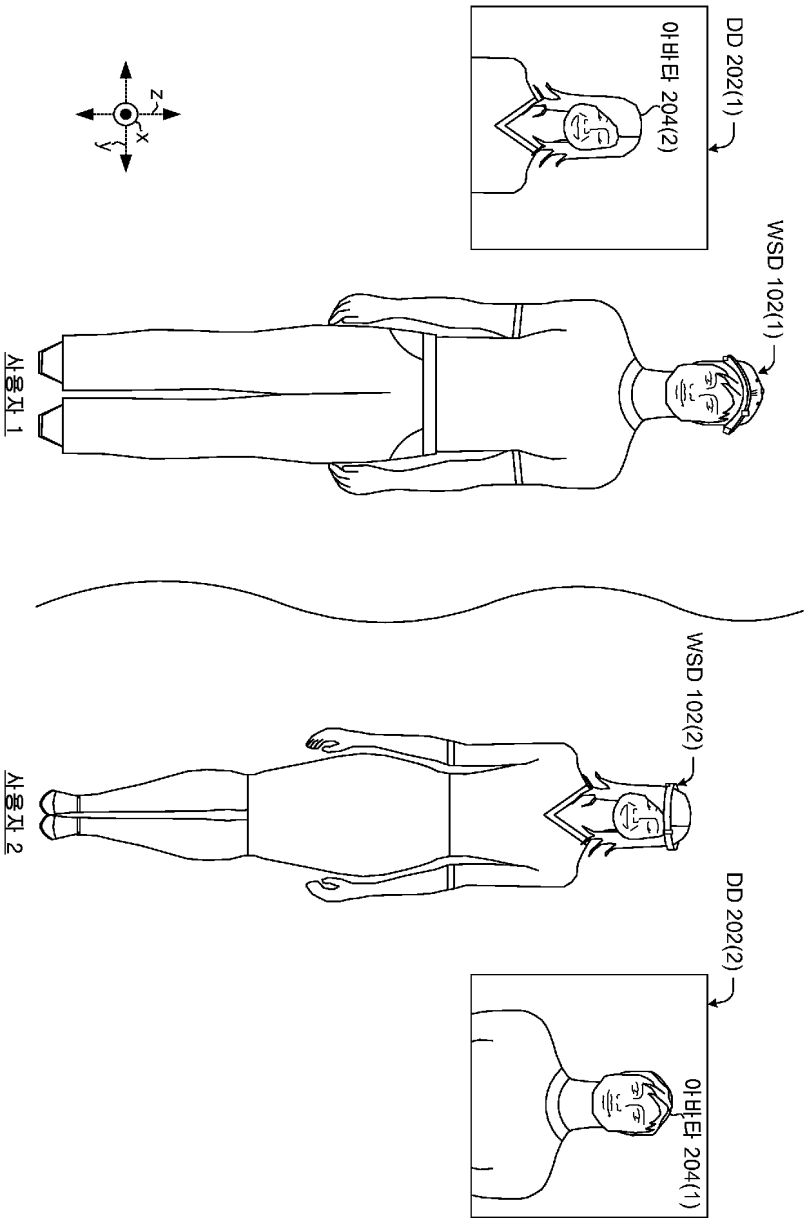
[0076] 시각화 정보에 관한 기술, 방법, 디바이스, 시스템 등을 구조적인 특징 및/또는 방법론적인 단계 특유의 언어로 주제에 대해 설명하였지만, 첨부하는 청구범위에서 규정되는 발명의 대상은 전술한 특징 또는 액트에 한정되어서는 안 된다고 이해된다. 그보다는, 특정 특징 및 액트는 청구하는 방법, 디바이스, 시스템 등을 구현하는 예시적인 형태로서 개시되어 있다.

도면

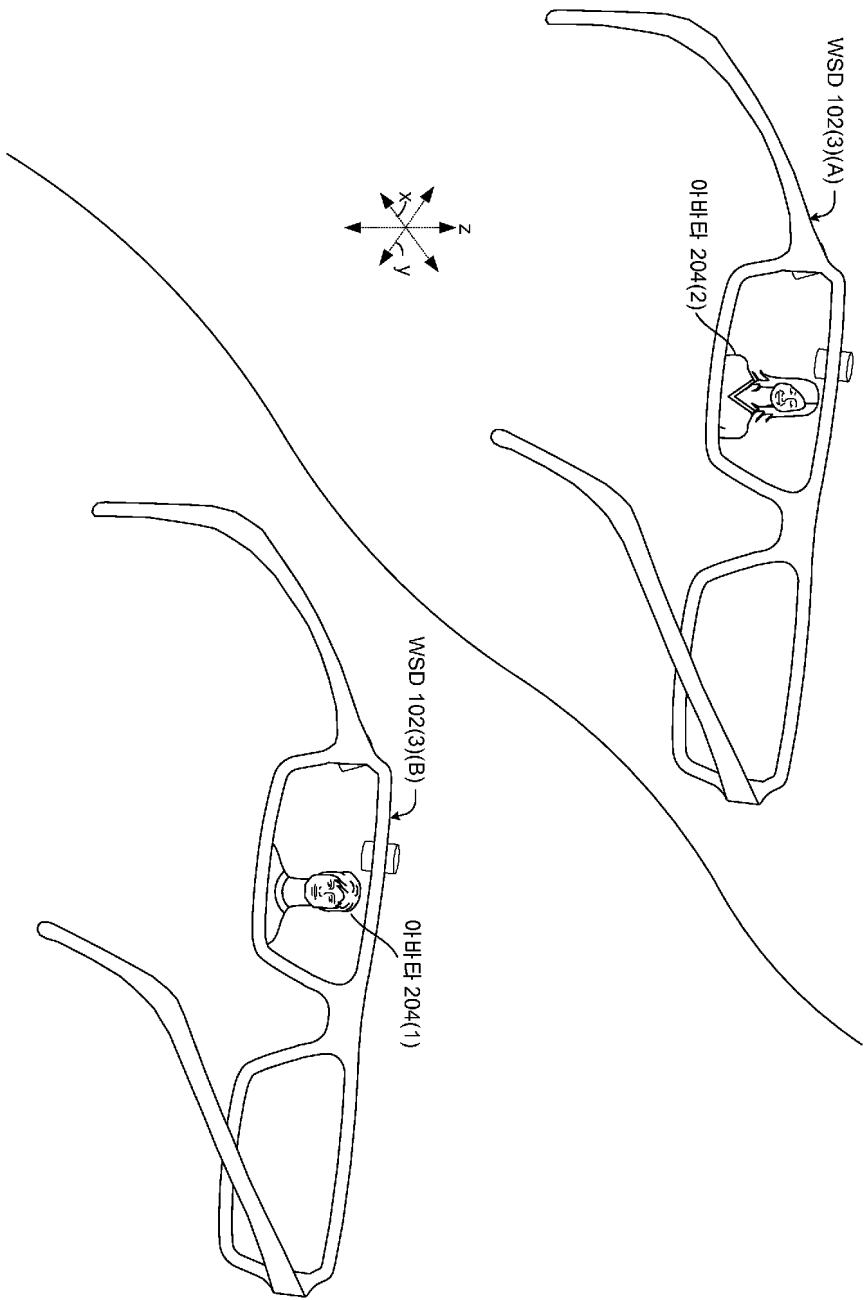
도면1



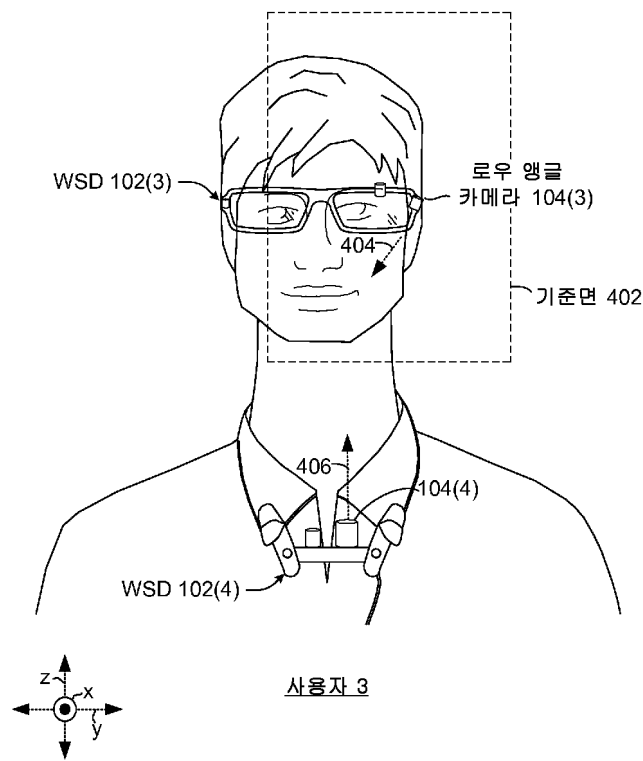
도면2



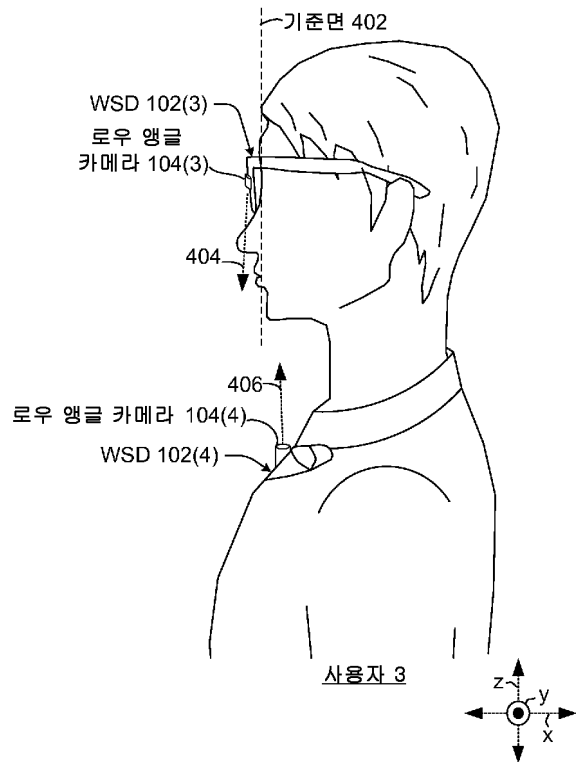
도면3



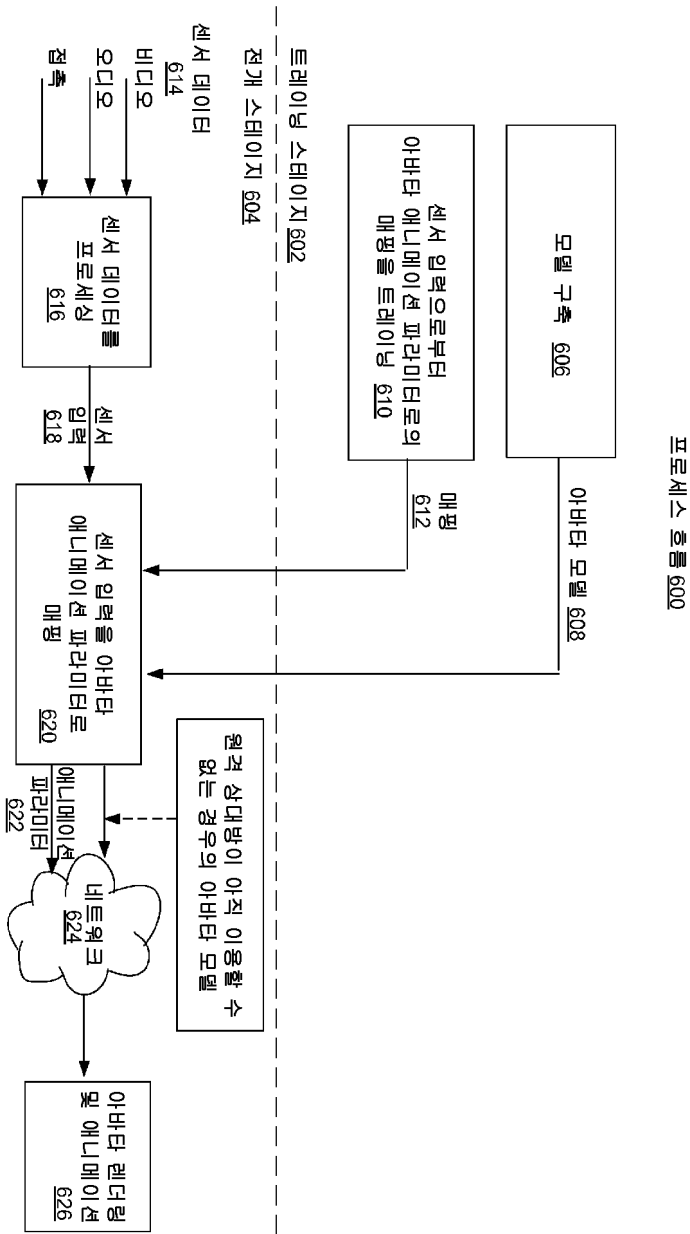
도면4



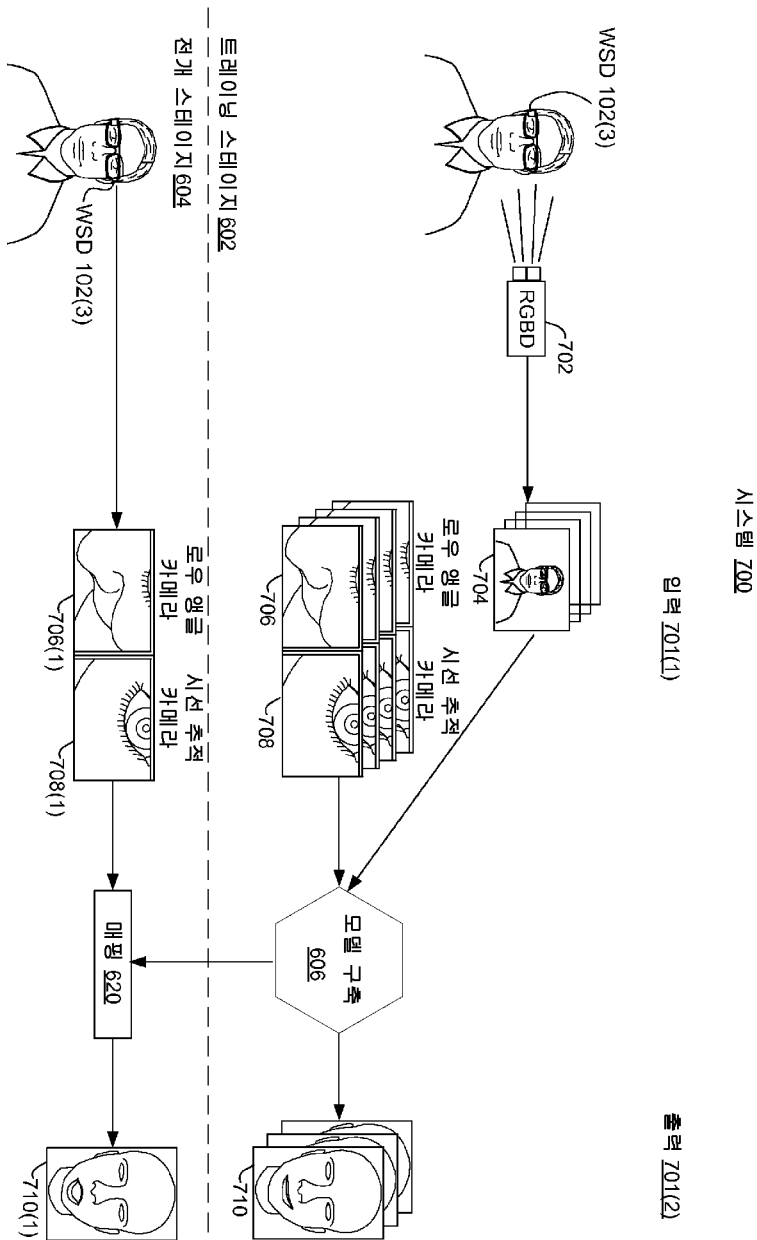
도면5



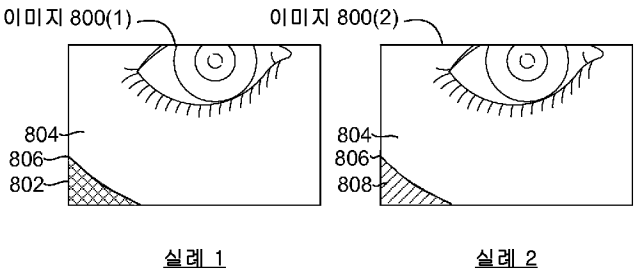
도면6



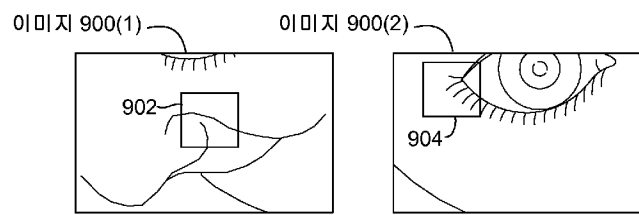
도면7



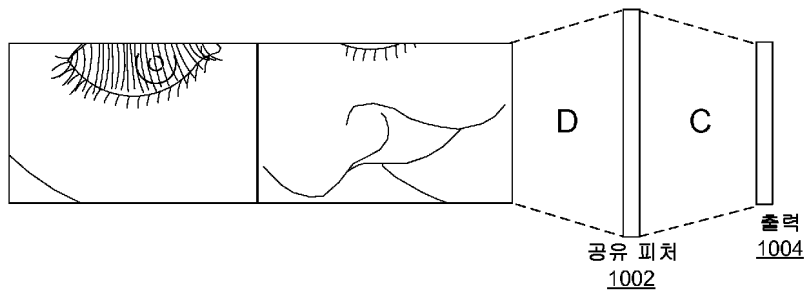
도면8



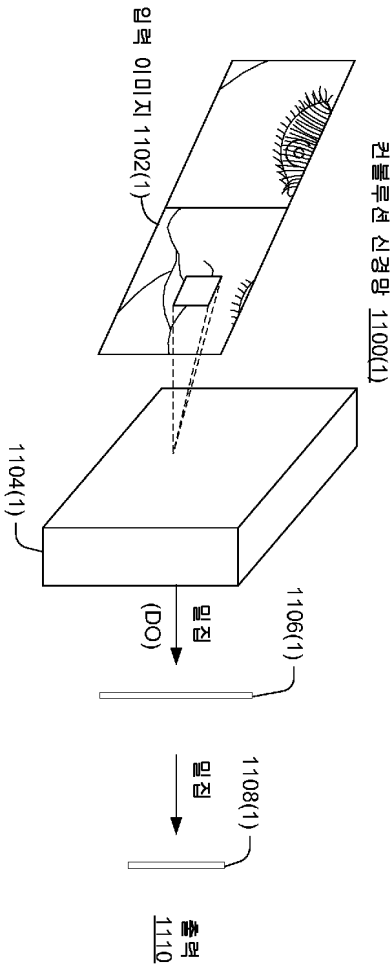
도면9



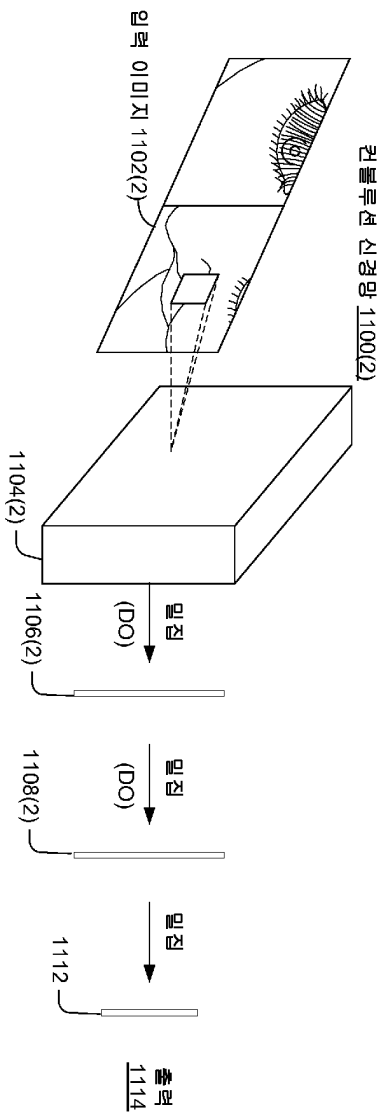
도면10



도면11a



도면11b



도면12

