(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0089525
(43) 공개일자 2024년06월20일

(51) 국제특허분류(Int. Cl.)
G06F 3/01 (2006.01) *G06F 3/16* (2018.01)
G06T 19/00 (2011.01) *G06V 20/20* (2022.01)
G06V 40/10 (2022.01) *G06V 40/20* (2022.01)
G10L 15/08 (2006.01)
(52) CPC특허분류
G06F 3/011 (2022.02)
G06F 3/017 (2013.01)
(21) 출원번호 10-2024-7015476
(22) 출원일자(국제) 2022년10월11일
심사청구일자 2024년05월09일
(85) 번역문제출일자 2024년05월09일
(86) 국제출원번호 PCT/US2022/046273
(87) 국제공개번호 WO 2023/064268
국제공개일자 2023년04월20일
(30) 우선권주장
17/498,510 2021년10월11일 미국(US)

(71) 출원인
스냅 인코포레이티드
미국 90405 캘리포니아주 산타 모니카 써티퍼스트 스트리트 3000
(72) 발명자
조하르, 마탄
미국 90405 캘리포니아 산타 모니카 31번 스트리트 3000
자오, 얀리
미국 90405 캘리포니아 산타 모니카 31번 스트리트 3000
(뒷면에 계속)
(74) 대리인
양영준, 김연송, 백만기

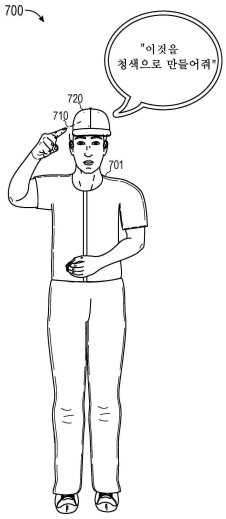
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **자세 및 음성 입력으로부터의 의도 추론**

(57) 요 약

본 개시내용의 양태들은 적어도 하나의 프로그램을 저장하는 컴퓨터 판독가능 저장 매체, 및 사람을 묘사하는 이미지를 수신하는 것, 사람의 골격 관절들의 세트를 식별하는 것, 및 골격 관절들의 세트의 위치확인에 기초하여 이미지에 묘사된 사람의 자세를 식별하는 것을 포함하는 동작들을 수행하기 위한 방법을 포함하는 시스템을 수반한다. 동작들은 또한 AR 동작을 수행하라는 요청 및 모호한 의도를 포함하는 음성 입력을 수신하는 것, 이미지에 묘사된 사람의 자세에 기초하여 음성 입력의 모호한 의도를 분별하는 것, 및 음성 입력을 수신하는 것에 응답하여, 이미지에 묘사된 사람의 자세에 기초하여 음성 입력의 모호한 의도를 분별하는 것에 기초하여 AR 동작을 수행하는 것을 포함한다.

대 표 도 - 도7



(52) CPC특허분류
G06F 3/167 (2013.01)
G06T 19/006 (2013.01)
G10L 15/08 (2013.01)
G06T 2210/16 (2013.01)
G10L 2015/088 (2013.01)
(72) 발명자
풀커슨, 브라이언
미국 90405 캘리포니아 산타 모니카 31번 스트리트 3000
베르거, 이타마르
미국 90405 캘리포니아 산타 모니카 31번 스트리트 3000

명 세 서

청구범위

청구항 1

방법은

하나 이상의 프로세서에 의해, 사람을 묘사하는 이미지를 수신하는 단계;

상기 사람의 골격 관절들의 세트를 식별하는 단계;

상기 골격 관절들의 세트의 위치확인에 기초하여 상기 이미지에 묘사된 상기 사람의 자세를 식별하는 단계;

증강 현실(AR) 동작을 수행하라는 요청을 포함하는 음성 입력을 수신하는 단계- 상기 음성 입력은 모호한 의도를 포함함 -;

상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 단계; 및

상기 음성 입력을 수신하는 것에 응답하여, 상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 것에 기초하여 상기 AR 동작을 수행하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서, 상기 모호한 의도는 대명사 또는 부사를 포함하고, 상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 대명사 또는 부사에 대응하는 타겟 객체 또는 방향을 식별하는 단계를 더 포함하는, 방법.

청구항 3

제1항 또는 제2항에 있어서,

상기 자세가 타겟 자세와 매칭된다고 결정하는 단계; 및

상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여, 상기 타겟 객체가 상기 이미지에 묘사된 사람을 포함한다고 결정하는 단계를 더 포함하는, 방법.

청구항 4

제3항에 있어서,

상기 음성 입력을 수신하는 것에 응답하여 그리고 상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여 상기 이미지에 묘사된 사람에게 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 자세가 타겟 자세와 매칭된다고 결정하는 단계; 및

상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여, 상기 타겟 객체가 상기 이미지에 묘사된 상기 사람의 신체 부분을 포함한다고 결정하는 단계를 더 포함하는, 방법.

청구항 6

제5항에 있어서,

상기 음성 입력을 수신하는 것에 응답하여 그리고 상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여 상기 이미지에 묘사된 상기 사람의 상기 신체 부분에 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 7

제5항 또는 제6항에 있어서,

상기 신체 부분에 대응하는 상기 이미지의 부분 상에 묘사된 현실 세계 패션 아이템을 식별하는 단계;

상기 이미지에 묘사된 상기 현실 세계 패션 아이템을 세그멘테이션하여 상기 현실 세계 패션 아이템의 세그멘테이션을 생성하는 단계; 및

상기 현실 세계 패션 아이템의 세그멘테이션에 기초하여, 상기 현실 세계 패션 아이템의 속성을 상기 하나 이상의 AR 요소로 대체하는 단계를 더 포함하는 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 사람은 제1 인물이고, 상기 이미지는 제2 인물을 묘사하고,

상기 자세가 타겟 자세와 매칭된다고 결정하는 단계; 및

상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여, 상기 모호한 의도의 대명사가 상기 이미지에 묘사된 상기 제2 인물을 지칭한다고 결정하는 단계를 더 포함하는, 방법.

청구항 9

제8항에 있어서,

상기 음성 입력을 수신하는 것에 응답하여 그리고 상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여 상기 이미지에 묘사된 상기 제2 인물에게 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 10

제8항 또는 제9항에 있어서,

상기 음성 입력이 상기 제2 인물이 착용한 패션 아이템을 지칭한다고 결정하는 단계;

상기 음성 입력을 수신하는 것에 응답하여 상기 패션 아이템의 세그멘테이션을 생성하기 위해 상기 제2 인물이 착용한 상기 패션 아이템을 세그멘테이션하는 단계; 및

상기 패션 아이템의 세그멘테이션에 기초하여 상기 제2 인물이 착용한 패션 아이템에 상기 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 11

제8항 내지 제10항 중 어느 한 항에 있어서,

상기 음성 입력을 수신하는 것에 응답하여 그리고 상기 자세가 상기 타겟 자세와 매칭된다고 결정하는 것에 응답하여 상기 이미지에 묘사된 상기 제2 인물이 착용한 패션 아이템의 속성에 기초하여 상기 이미지에 묘사된 상기 제1 인물에게 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 12

제8항 내지 제11항 중 어느 한 항에 있어서,

상기 음성 입력이 상기 패션 아이템을 지칭한다고 결정하는 단계;

상기 음성 입력을 수신하는 것에 응답하여 상기 패션 아이템의 제1 세그멘테이션을 생성하기 위해 상기 제2 인물이 착용한 상기 패션 아이템을 세그멘테이션하는 단계;

상기 음성 입력을 수신하는 것에 응답하여 상기 제1 인물이 착용한 상기 패션 아이템을 세그멘테이션하여 상기 패션 아이템의 제2 세그멘테이션을 생성하는 단계;

상기 제1 세그멘테이션에 기초하여 상기 제2 인물이 착용한 상기 패션 아이템의 픽셀 값들을 획득하는 단계; 및

상기 패션 아이템의 상기 제2 세그멘테이션 및 상기 획득된 픽셀 값들에 기초하여 상기 제1 인물이 착용한 상기 패션 아이템에 상기 하나 이상의 AR 요소를 적용하는 단계를 더 포함하는, 방법.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서,

상기 자세가 타겟 자세와 매칭된다고 결정하는 단계; 및

자세가 타겟 자세와 매칭된다고 결정하는 것에 응답하여, 모호한 의도의 대명사가 이미지에 묘사된 인물이 착용한 패션 아이템의 속성을 지칭한다고 결정하는 단계를 더 포함하는, 방법.

청구항 14

제13항에 있어서,

상기 타겟 자세가 가리키는 상기 이미지의 부분에서의 픽셀 값을 결정하는 단계; 및

상기 결정된 픽셀 값에 기초하여 상기 이미지에 묘사된 상기 인물이 착용한 상기 패션 아이템의 속성을 대체하는 단계를 더 포함하는, 방법.

청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서, 상기 자세는 특정 방향을 따라 가리키는 신체 부분을 포함하고,

상기 특정 방향을 따라 상기 신체 부분으로부터 연장되는 라인과 교차하는 객체를 식별하는 단계; 및

음성 입력의 모호한 의도가 신체 부분으로부터 연장되는 라인과 교차하는 객체에 기초하여 식별된 객체를 지칭함을 분별하는 단계를 더 포함하는, 방법.

청구항 16

제15항에 있어서, 상기 객체는 상기 이미지에 묘사된 상기 인물이 착용한 제1 패션 아이템, 상기 이미지에 묘사된 제2 인물이 착용한 제2 패션 아이템, 상기 제2 인물, 또는 상기 이미지에서의 픽셀의 색상 중 적어도 하나를 포함하는, 방법.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 AR 동작을 수행하라는 상기 요청은 상기 이미지에 묘사된 상기 사람에게 패션 아이템을 추가하는 것 또는 상기 이미지에 묘사된 상기 인물이 착용한 주어진 패션 아이템의 색상 또는 속성을 변경하는 것 중 적어도 하나를 포함하는, 방법.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, 상기 자세는 형상을 표현하고, 상기 모호한 의도의 대명사는 상기 자세에 의해 표현되는 형상에 기초하여 해석되고, 상기 음성 입력을 수신하는 것에 응답하여 상기 자세에 의해 표현되는 형상에 기초하여 하나 이상의 AR 요소를 제시하는 단계를 더 포함하는, 방법.

청구항 19

시스템으로서,

동작들을 수행하도록 구성된 하나 이상의 프로세서를 포함하고, 상기 동작들은

사람을 묘사하는 이미지를 수신하는 동작;

상기 사람의 골격 관절들의 세트를 식별하는 동작;

상기 골격 관절들의 세트의 위치확인에 기초하여 상기 이미지에 묘사된 상기 사람의 자세를 식별하는 동작;

증강 현실(AR) 동작을 수행하라는 요청을 포함하는 음성 입력을 수신하는 동작 -상기 음성 입력은 모호한 의도를 포함함 -;

상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 동작; 및

상기 음성 입력을 수신하는 것에 응답하여, 상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 것에 기초하여 상기 AR 동작을 수행하는 동작을 포함하는, 시스템.

청구항 20

머신의 하나 이상의 프로세서에 의해 실행될 때, 상기 머신으로 하여금 동작들을 수행하게 하는 명령어들을 포함하는 증강 현실 시스템을 포함하는 비일시적 머신 판독가능 저장 매체로서, 상기 동작들은

사람을 묘사하는 이미지를 수신하는 동작;

상기 사람의 골격 관절들의 세트를 식별하는 동작;

상기 골격 관절들의 세트의 위치확인에 기초하여 상기 이미지에 묘사된 상기 사람의 자세를 식별하는 동작;

증강 현실(AR) 동작을 수행하라는 요청을 포함하는 음성 입력을 수신하는 동작- 상기 음성 입력은 모호한 의도를 포함함 -;

상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 동작; 및

상기 음성 입력을 수신하는 것에 응답하여, 상기 이미지에 묘사된 상기 사람의 자세에 기초하여 상기 음성 입력의 모호한 의도를 분별하는 것에 기초하여 상기 AR 동작을 수행하는 동작을 포함하는, 비일시적 머신 판독가능 저장 매체.

발명의 설명

기 술 분 야

[0001] 우선권 주장

[0002] 본 출원은 2021년 10월 11일에 출원된 미국 특허 출원 일련 번호 17/498,510에 대한 우선권의 이익을 주장하며, 이는 그 전문이 본원에 참조로 포함된다.

[0003] 기술 분야

[0004] 본 개시내용은 일반적으로 메시징 애플리케이션을 사용하여 증강 현실 경험들을 제공하는 것에 관한 것이다.

배 경 기 술

[0005] AR(Augmented-Reality)은 가상 환경의 수정이다. 예를 들어, VR(Virtual Reality)에서, 사용자는 가상 세계에 완전히 몰입되는 반면, AR에서, 사용자는 가상 객체들이 현실 세계에 조합 또는 중첩되는 세계에 몰입된다. AR 시스템은 현실 세계 환경과 그리고 서로 현실적으로 상호작용하는 가상 객체들을 생성하고 제시하는 것을 목표로 한다. AR 응용의 예는 단일 또는 다중 플레이어 비디오 게임, 인스턴트 메시징 시스템 등을 포함할 수 있다.

도면의 간단한 설명

[0006] 반드시 축척에 따라 그려지지는 않은 도면에서, 유사한 번호는 상이한 도면에서 유사한 컴포넌트를 설명할 수 있다. 임의의 특정 요소 또는 동작에 대한 논의를 용이하게 식별할 수 있도록, 참조 번호의 최상위 숫자 또는 숫자들은 해당 요소가 최초 도입되는 도면 번호를 참조한다. 일부 비제한적인 예들이 첨부 도면들의 도면들에 예시된다.

도 1은 일부 예들에 따른, 본 개시내용이 배치될 수 있는 네트워크화된 환경의 도식적 표현이다.

도 2는 일부 예시적인 실시예들에 따른, 메시징 클라이언트 애플리케이션의 도식적 표현이다.

도 3은 일부 예들에 따른, 데이터베이스에 유지되는 데이터 구조의 도식적 표현이다.

도 4는 일부 예시적인 실시예들에 따른, 메시지의 도식적 표현이다.

도 5는 예시적인 예들에 따른 예시적인 의도 결정 시스템을 도시하는 블록도이다.

도 6, 도 7, 및 도 8은 일부 예들에 따른, 의도 결정 시스템의 출력들의 도식적 표현들이다.

도 9는 일부 예들에 따른, 의도 결정 시스템의 예시적인 동작들을 예시하는 흐름도이다.

도 10은 일부 예들에 따른, 머신으로 하여금 본 명세서에서 논의된 방법론들 중 어느 하나 이상을 수행하게 하

기 위해 명령어들의 세트가 실행될 수 있는 컴퓨터 시스템 형태의 머신의 도식적 표현이다.

도 11은 예들이 구현될 수 있는 소프트웨어 아키텍처를 도시하는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0007] 이하의 설명은 본 개시내용의 예시적인 예들을 구현하는 시스템들, 방법들, 기술들, 명령어 시퀀스들, 및 컴퓨팅 머신 프로그램 제품들을 포함한다. 이하의 설명에서, 설명의 목적상, 다양한 예들에 대한 이해를 제공하기 위해 다수의 구체적인 세부사항들이 기재되어 있다. 그러나, 관련 기술분야의 통상의 기술자들은 예들이 이러한 특정 세부사항들 없이 실시될 수 있다는 것을 명백히 알 수 있을 것이다. 일반적으로, 잘 알려진 명령어 인스턴스, 프로토콜, 구조 및 기술은 반드시 상세히 나타나 있는 것은 아니다.

[0008] 전형적으로, VR 및 AR 시스템들은 사용자의 이미지를 캡처하고, 추가로, 깊이 센서를 사용하여 이미지에 묘사된 현실 세계 인체의 깊이 맵을 획득함으로써 주어진 사용자를 표현하는 이미지들을 디스플레이한다. 깊이 맵과 이미지를 함께 처리함으로써, VR 및 AR 시스템들은 이미지에서의 사용자의 위치를 검출할 수 있고, 이미지들에서의 사용자 또는 배경을 적절하게 수정할 수 있다. 이러한 시스템들은 잘 작동하지만, 깊이 센서에 대한 필요성은 그들의 응용들의 범위를 제한한다. 이는 이미지들을 수정할 목적으로 사용자 디바이스들에 깊이 센서들을 추가하는 것이 디바이스들의 전체 비용 및 복잡성을 증가시켜, 디바이스들을 덜 매력적으로 만들기 때문이다.

[0009] 특정 시스템들은 이미지들을 수정하기 위해 깊이 센서들을 사용할 필요성을 제거한다. 예를 들어, 특정 시스템들은 사용자의 얼굴이 검출되는 화상 회의에서 사용자들이 배경을 대체하는 것을 허용한다. 구체적으로, 이러한 시스템들은 사용자의 얼굴을 묘사하는 이미지들에서 배경을 식별하기 위해 사용자의 얼굴을 인식하도록 최적화된 특수화된 기술들을 사용할 수 있다. 그 다음, 이들 시스템들은 배경을 묘사하는 픽셀들만을 대체할 수 있어서, 현실 세계 배경이 이미지들에서 대안적인 배경으로 대체된다. 그러나, 이러한 시스템들은 일반적으로 사용자의 전신을 인식할 수 없다. 이 때문에, 사용자가 카메라로부터 임계 거리를 초과하고, 그래서, 단지 사용자의 얼굴만이 아니라 더 많은 것이 카메라에 의해 캡처되는 경우, 배경을 대안적인 배경으로 대체하는 것에 실패하기 시작한다. 이러한 경우들에서, 이미지 품질은 심각하게 영향을 받고, 사용자의 얼굴 및 신체의 부분들이 시스템에 의해 부주의하게 제거될 수 있으며, 그 이유는 시스템이 이러한 부분들을 이미지들의 전경이 아니라 배경에 속하는 것으로 잘못 식별하기 때문이다. 또한, 이러한 시스템들은 한명보다 더 많은 사용자가 이미지 또는 비디오 피드에 묘사될 때 배경을 적절히 대체하는 데 실패한다. 이러한 시스템들은 일반적으로 이미지에서의 사용자의 전신을 배경과 구별할 수 없기 때문에, 이러한 시스템들은 또한 신체 부분(예를 들어, 얼굴)을 AR 그래픽으로 변환, 블렌딩, 변환, 변경 또는 모핑하는 것과 같은 시각적 효과들을 사용자의 신체의 특정 부분들에 적용할 수 없다.

[0010] 일부 AR 시스템들은 사용자들이 음성을 사용하여 다양한 AR 동작들을 수행하는 것을 허용한다. 즉, AR 시스템들은 음성 또는 음성 입력을 적절한 커맨드들로 변환하기 위해 음성 인식 엔진들을 이용한다. 이러한 시스템들은 일반적으로 잘 작동하지만, 사용자들이 음성 입력을 제공할 때 엄격한 형식 지침들을 따를 것을 요구한다. 수신되는 음성 입력이 지침들을 준수하지 못하면, AR 시스템들은 올바른 AR 동작들을 수행하지 못하거나 적절한 커맨드들을 인식하지 못한다. 예를 들어, 수신되는 음성이 하나 이상의 대명사 또는 부사를 포함하는 경우, AR 시스템은 AR 시스템이 대명사들 또는 부사들의 의도를 명확하게 할 수 없기 때문에 수행할 정확한 동작을 식별할 수 없다. 이로 인해 사용자들은 그들의 관심들을 자유롭게 말할 수 없고, 음성 입력을 사용하여 AR 시스템들을 제어하는 즐거움으로부터 멀어진다. 그 결과, 음성 인식을 수행하는 데 전용되는 리소스들이 낭비된다.

[0011] 개시된 기술들은 이미지에 묘사된 사람에 대한 골격 추적을 수행하고 골격 추적과 연관된 자세를 결정함으로써 전자 디바이스를 사용하는 효율을 개선한다. 개시된 기술들은 하나 이상의 대명사 또는 부사와 같은 모호한 의도를 포함하고 AR 동작의 수행을 요청하는 음성 입력을 수신할 수 있다. 개시된 기술들은 음성 입력의 의도를 식별하고 명확하게 하기 위해(예를 들어, 대명사들 또는 부사들에 의해 표현되는 객체, 아이템, 대상, 방향 또는 수량을 식별하기 위해) 이미지에 묘사된 사람과 연관된 자세를 활용할 수 있다. 음성 입력의 의도를 명확히 한 후에, 개시된 기술들은 요청된 AR 동작을 수행할 수 있다. 이것은 사용자가 그들의 음성을 사용하여 AR 시스템들을 제어하는 것을 허용함으로써 AR 시스템들의 동작을 상당히 개선한다. 또한, 사용자들이 엄격한 음성 제어 지침들을 학습하고 따를 필요가 없기 때문에, 개시된 AR 시스템을 사용하는 학습 굴곡이 상당히 감소되고, 이는 더 많은 사용자들이 음성 제어를 사용하여 개시된 AR 시스템을 동작시킬 수 있게 하며, 이는 리소스들이 낭비되는 것을 피하고 디바이스의 전체 기능을 개선한다.

[0012] 즉, 개시된 기술들은 음성에 기초하여 이미지 또는 비디오에 AR 그래픽들을 추가하는 프로세스를 단순화하며,

이는 이러한 AR 그래픽들을 생성하는 데 있어서 설계 제약들 및 비용들을 상당히 감소시키고 처리 복잡성들 및 전력 및 메모리 요건들의 양을 감소시킨다. 이것은 또한 AR 그래픽이 객체를 묘사하는 이미지 또는 비디오에 묘사된 현실 세계 환경의 일부인 착각을 개선한다. 이는 소규모 모바일 디바이스들 상에서 실시간으로 기본 이미지 또는 비디오에 AR 그래픽들의 매끄러운, 그리고, 효율적인 추가를 가능하게 한다. 개시된 기술들은 모바일 디바이스가 이미지들/비디오들을 서버에 전송할 필요 없이 전적으로 또는 주로 모바일 디바이스 상에서 적용될 수 있다. 다른 예들에서, 개시된 기술들은 전적으로 또는 주로 원격 서버 상에서 적용되거나 모바일 디바이스와 서버 사이에 분할될 수 있다.

[0013] 그 결과, 사용자가 그들의 음성과 상호작용하고 그들의 음성을 사용하여 선택하는 것이 직관적인 방식으로, 사람이 신체 형상, 신체 상태, 신체 특성들, 위치, 및 회전에 대한 변화들을 포함하는 3차원(3D)으로 비디오 주위를 이동하는 동안 이미지 또는 비디오에 AR 그래픽을 추가하는 현실적인 디스플레이가 제공될 수 있다. 본 명세서에서 사용될 때, "의류 물품", "패션 아이템", 및 "의복"은 상호교환가능하게 사용되고, 동일한 의미를 갖는 것으로 이해되어야 한다. 이는 전자 디바이스를 사용함에 있어서 사용자의 전체 경험을 개선한다. 또한, 깊이 센서를 사용하지 않고 이러한 AR 경험들을 제공함으로써, 작업 달성에 필요한 시스템 리소스들의 전체 양이 감소된다.

[0014] 네트워크화된 컴퓨팅 환경

[0015] 도 1은 네트워크를 통해 데이터(예를 들어, 메시지들 및 연관된 콘텐츠)를 교환하기 위한 예시적인 메시징 시스템(100)을 도시하는 블록도이다. 메시징 시스템(100)은 클라이언트 디바이스(102)의 다수의 인스턴스들을 포함하고, 이들 각각은 메시징 클라이언트(104) 및 다른 외부 애플리케이션들(109)(예를 들어, 제3자 애플리케이션들)을 포함하는 다수의 애플리케이션들을 호스팅한다. 각각의 메시징 클라이언트(104)는 네트워크(112)(예를 들어, 인터넷)를 통해 메시징 클라이언트(104)의 다른 인스턴스들(예를 들어, 각각의 다른 클라이언트 디바이스들(102) 상에 호스팅됨), 메시징 서버 시스템(108) 및 외부 앱(들) 서버들(110)에 통신가능하게 결합된다. 메시징 클라이언트(104)는 또한 애플리케이션 프로그래밍 인터페이스(API)들을 사용하여 로컬 호스팅된 제3자 애플리케이션들, 예컨대, 외부 앱들(109)과 통신할 수 있다.

[0016] 메시징 클라이언트(104)는 네트워크(112)를 통해 다른 메시징 클라이언트들(104)과 그리고 메시징 서버 시스템(108)과 통신하고 데이터를 교환할 수 있다. 메시징 클라이언트들(104) 사이에, 그리고 메시징 클라이언트(104)와 메시징 서버 시스템(108) 사이에 교환되는 데이터는 기능들(예를 들어, 기능들을 기동하기 위한 커맨드들)뿐만 아니라 페이로드 데이터(예를 들어, 텍스트, 오디오, 비디오 또는 다른 멀티미디어 데이터)를 포함한다.

[0017] 메시징 서버 시스템(108)은 네트워크(112)를 통해 특정 메시징 클라이언트(104)에 서버측 기능성을 제공한다. 메시징 시스템(100)의 특정 기능들이 메시징 클라이언트(104)에 의해 또는 메시징 서버 시스템(108)에 의해 수행되는 것으로 본 명세서에 설명되지만, 메시징 클라이언트(104) 또는 메시징 서버 시스템(108) 내의 특정 기능성의 위치는 설계 선택일 수 있다. 예를 들어, 초기에는 메시징 서버 시스템(108) 내에 특정 기술 및 기능성을 배치하지만, 나중에 클라이언트 디바이스(102)가 충분한 처리 용량을 갖는 메시징 클라이언트(104)로 이 기술 및 기능성을 이주시키는 것이 기술적으로 바람직할 수 있다.

[0018] 메시징 서버 시스템(108)은 메시징 클라이언트(104)에 제공되는 다양한 서비스들 및 동작들을 지원한다. 이러한 동작들은 메시징 클라이언트(104)에 데이터를 송신하는 것, 그로부터 데이터를 수신하는 것, 및 그에 의해 생성된 데이터를 처리하는 것을 포함한다. 이 데이터는, 예들로서, 메시지 콘텐츠, 클라이언트 디바이스 정보, 지리위치 정보, 미디어 증강 및 오버레이, 메시지 콘텐츠 지속 조건, 소셜 네트워크 정보, 및 라이브 이벤트 정보를 포함할 수 있다. 메시징 시스템(100) 내의 데이터 교환들은 메시징 클라이언트(104)의 사용자 인터페이스들을 통해 이용가능한 기능들을 통해 기동되고 제어된다.

[0019] 이제 구체적으로 메시징 서버 시스템(108)을 참조하면, API 서버(116)가 애플리케이션 서버들(114)에 결합되어, 프로그램 방식의 인터페이스를 이에 제공한다. 애플리케이션 서버들(114)은 데이터베이스 서버(120)에 통신가능하게 결합되며, 이는 애플리케이션 서버들(114)에 의해 처리되는 메시지들과 연관된 데이터를 저장하는 데이터베이스(126)에 대한 액세스를 용이하게 한다. 유사하게, 웹 서버(128)는 애플리케이션 서버들(114)에 결합되고, 웹 기반 인터페이스들을 애플리케이션 서버들(114)에 제공한다. 이를 위해, 웹 서버(128)는 하이퍼텍스트 전송 프로토콜(Hypertext Transfer Protocol, HTTP) 및 몇몇 다른 관련 프로토콜을 통해 착신 네트워크 요청들을 처리한다.

[0020] API 서버(116)는 클라이언트 디바이스(102)와 애플리케이션 서버들(114) 사이에서 메시지 데이터(예를 들어, 커맨드들 및 메시지 페이로드들)를 수신 및 송신한다. 구체적으로, API 서버(116)는 애플리케이션 서버들(114)의 기능성을 기동하기 위해 메시징 클라이언트(104)에 의해 호출되거나 질의될 수 있는 인터페이스들(예를 들어, 루틴들 및 프로토콜들)의 세트를 제공한다. API 서버(116)는 계정 등록, 로그인 기능성, 특정 메시징 클라이언트(104)로부터 다른 메시징 클라이언트(104)로의, 애플리케이션 서버들(114)을 통한 메시지들의 전송, 메시징 클라이언트(104)로부터 메시징 서버(118)로의 미디어 파일들(예를 들어, 이미지들 또는 비디오)의 전송, 및 다른 메시징 클라이언트(104)에 의한 가능한 액세스를 위해, 미디어 데이터의 컬렉션(예를 들어, 스토리)의 설정들, 클라이언트 디바이스(102)의 사용자의 친구들의 리스트의 검색, 이러한 컬렉션들의 검색, 메시지들 및 콘텐츠의 검색, 엔티티 그래프(예를 들어, 소셜 그래프)에 대한 엔티티들(예를 들어, 친구들)의 추가 및 삭제, 소셜 그래프 내의 친구들의 위치, 및 (예를 들어, 메시징 클라이언트(104)에 관련된) 애플리케이션 이벤트의 개최를 비롯하여, 애플리케이션 서버들(114)에 의해 지원되는 다양한 기능들을 노출시킨다.

[0021] 애플리케이션 서버들(114)은, 예를 들어, 메시징 서버(118), 이미지 처리 서버(122), 및 소셜 네트워크 서버(124)를 포함하는, 다수의 서버 애플리케이션들 및 서브시스템들을 호스팅한다. 메시징 서버(118)는, 특히 메시징 클라이언트(104)의 다수의 인스턴스들로부터 수신된 메시지들에 포함된 콘텐츠(예를 들어, 텍스트 및 멀티미디어 콘텐츠)의 집성 및 다른 처리와 관련된 다수의 메시지 처리 기술들 및 기능들을 구현한다. 더 상세히 설명되는 바와 같이, 다수의 소스로부터의 텍스트 및 미디어 콘텐츠는, 콘텐츠의 컬렉션들(예를 들어, 스토리 또는 갤러리라고 지칭됨)로 집성될 수 있다. 다음으로, 이러한 컬렉션들은 메시징 클라이언트(104)에 이용가능하게 된다. 데이터의 다른 프로세서- 및 메모리-집약적 처리는 또한, 이러한 처리를 위한 하드웨어 요건들을 고려하여, 메시징 서버(118)에 의해 서버측에서 수행될 수 있다.

[0022] 애플리케이션 서버들(114)은, 전형적으로 메시징 서버(118)로부터 전송되거나 그것에서 수신되는 메시지의 페이로드 내의 이미지들 또는 비디오에 대해, 다양한 이미지 처리 동작들을 수행하는 데 전용되는 이미지 처리 서버(122)를 또한 포함한다.

[0023] 이미지 처리 서버(122)는 증강 시스템(208)(도 2에 도시됨)의 스캔 기능성을 구현하는 데 사용된다. 스캔 기능성은 이미지가 클라이언트 디바이스(102)에 의해 캡처될 때 클라이언트 디바이스(102) 상에서 하나 이상의 AR 경험을 활성화하고 제공하는 것을 포함한다. 구체적으로, 클라이언트 디바이스(102) 상의 메시징 클라이언트(104)는 카메라를 활성화하기 위해 사용될 수 있다. 카메라는 하나 이상의 AR 경험의 하나 이상의 아이콘 또는 식별자와 함께 하나 이상의 실시간 이미지 또는 비디오를 사용자에게 디스플레이한다. 사용자는 대응하는 AR 경험을 론칭하거나 원하는 이미지 수정을 수행하기 위해 식별자들 중 주어진 하나를 선택할 수 있다(예를 들어, 비디오에서 사용자가 착용하고 있는 의복을 대체 또는 변경하거나, 또는 사람 또는 사용자의 신체의 일부를 AR 웨어울프 또는 AR 배트와 같은 AR 그래픽으로 모핑, 변경, 블렌딩 또는 변환함).

[0024] 일부 경우들에서, 사용자는 음성 입력을 제공함으로써 주어진 AR 경험을 활성화하기 위한 커맨드를 말할 수 있다. 클라이언트 디바이스(102)는 음성 입력을 수신할 수 있는 마이크로폰을 포함한다. 클라이언트 디바이스(102)는 음성 입력을 처리하고, 예컨대 하나 이상의 대명사 또는 부사를 식별함으로써, 음성 입력의 의도에서의 모호성을 검출할 수 있다. 클라이언트 디바이스(102)는 이미지 또는 비디오에 묘사된 사용자에 대한 신체 자세 정보를 수신할 수 있고 신체 자세를 사용하여 의도를 식별하거나 명확하게 할 수 있다. 즉, 클라이언트 디바이스(102)는 음성 입력에서 대명사 또는 부사가 지칭하는 정확한 객체, 아이템, 장소, 또는 방향을 식별함으로써 음성 입력에서 요청된 AR 동작을 수행할 수 있다.

[0025] 소셜 네트워크 서버(124)는 다양한 소셜 네트워크화 기능들 및 서비스들을 지원하고 이러한 기능들 및 서비스들을 메시징 서버(118)에 이용가능하게 한다. 이를 위해, 소셜 네트워크 서버(124)는 데이터베이스(126) 내에 (도 3에 도시된 바와 같은) 엔티티 그래프(308)를 유지하고 액세스한다. 소셜 네트워크 서버(124)에 의해 지원되는 기능들 및 서비스들의 예들은 특정 사용자가 관계들을 갖거나 "팔로우하고 있는" 메시징 시스템(100)의 다른 사용자들의 식별, 및 또한 특정 사용자의 다른 엔티티들 및 관심들의 식별을 포함한다.

[0026] 메시징 클라이언트(104)로 돌아가서, 외부 리소스(예를 들어, 제3자 애플리케이션(109) 또는 애플릿)의 특징들 및 기능들은 메시징 클라이언트(104)의 인터페이스를 통해 사용자에게 이용가능하게 된다. 메시징 클라이언트(104)는 외부 앱들(109)과 같은 외부 리소스(예를 들어, 제3자 리소스)의 특징들을 론칭하거나 액세스하기 위한 옵션의 사용자 선택을 수신한다. 외부 리소스는 클라이언트 디바이스(102) 상에 설치된 제3자 애플리케이션(외부 앱들(109))(예를 들어, "네이티브 앱"), 또는 클라이언트 디바이스(102) 상에서 또는 클라이언트 디바이스(102)의 원격에서(예를 들어, 제3자 서버들(110) 상에서) 호스팅되는 제3자 애플리케이션의 소규모 버전(예를

들어, "애플릿")일 수 있다. 제3자 애플리케이션의 소규모 버전은 제3자 애플리케이션의 특징들 및 기능들의 서브세트(예를 들어, 제3자 독립형 애플리케이션의 풀-스케일, 네이티브 버전)를 포함하고, 마크업-언어 문서를 사용하여 구현된다. 일 예에서, 제3자 애플리케이션의 소규모 버전(예를 들어, "애플릿")은 제3자 애플리케이션의 웹 기반 마크업 언어 버전이고 메시징 클라이언트(104)에 임베딩된다. 마크업 언어 문서들(예를 들면, .*ml 파일)을 사용하는 것 외에도, 애플릿은 스크립팅 언어(예를 들면, .*js 파일 또는 .json 파일) 및 스타일 시트(예를 들면, .*ss 파일)를 포함할 수 있다.

[0027] 외부 리소스(외부 앱(109))의 특징들을 론칭하거나 액세스하기 위한 옵션의 사용자 선택을 수신하는 것에 응답하여, 메시징 클라이언트(104)는 선택된 외부 리소스가 웹 기반 외부 리소스인지 또는 로컬로 설치된 외부 애플리케이션인지를 결정한다. 일부 경우들에서, 클라이언트 디바이스(102) 상에 로컬로 설치되는 외부 애플리케이션들(109)은 메시징 클라이언트(104)와 독립적으로 그리고 별개로, 예컨대, 클라이언트 디바이스(102)의 홈 스크린 상에서, 외부 애플리케이션(109)에 대응하는 아이콘을 선택하는 것에 의해서, 론칭될 수 있다. 이러한 외부 애플리케이션들의 소규모 버전들은 메시징 클라이언트(104)를 통해 론칭되거나 액세스될 수 있고, 일부 예들에서, 소규모 외부 애플리케이션의 어떠한 부분도 메시징 클라이언트(104) 외부에서 액세스되지 않거나 제한된 부분들이 액세스될 수 있다. 소규모 외부 애플리케이션은 메시징 클라이언트(104)가 외부 앱(들) 서버(110)로부터 소규모 외부 애플리케이션과 연관된 마크업 언어 문서를 수신하고 이러한 문서를 처리함으로써 론칭될 수 있다.

[0028] 외부 리소스가 로컬로 설치된 외부 애플리케이션(109)이라고 결정하는 것에 응답하여, 메시징 클라이언트(104)는 외부 애플리케이션(109)에 대응하는 로컬로 저장된 코드를 실행함으로써 외부 애플리케이션(109)을 론칭하도록 클라이언트 디바이스(102)에 지시한다. 외부 리소스가 웹 기반 리소스라고 결정하는 것에 응답하여, 메시징 클라이언트(104)는 선택된 리소스에 대응하는 마크업 언어 문서를 획득하기 위해 외부 앱(들) 서버들(110)과 통신한다. 그 다음, 메시징 클라이언트(104)는 획득된 마크업 언어 문서를 처리하여 메시징 클라이언트(104)의 사용자 인터페이스 내에 웹 기반 외부 리소스를 제시한다.

[0029] 메시징 클라이언트(104)는 클라이언트 디바이스(102)의 사용자, 또는 이러한 사용자와 관련된 다른 사용자들(예를 들어, "친구들")에게 하나 이상의 외부 리소스에서 발생하는 활동을 통지할 수 있다. 예를 들어, 메시징 클라이언트(104)는 메시징 클라이언트(104)에서의 대화(예를 들어, 채팅 세션)의 참가자들에게 사용자들의 그룹의 하나 이상의 멤버에 의한 외부 리소스의 현재 또는 최근 사용에 관한 통지들을 제공할 수 있다. 하나 이상의 사용자들은 활성 외부 리소스에 참여하거나 (친구들의 그룹에서) 최근에 사용되었지만 현재 비활성인 외부 리소스를 론칭하도록 초대될 수 있다. 외부 리소스는 대화의 참가자들에게, 각각이 각각의 메시징 클라이언트(104)를 사용하여, 외부 리소스에서의 아이템, 스테이터스, 상태, 또는 위치를 사용자들의 그룹의 하나 이상의 멤버와 채팅 세션으로 공유하는 능력을 제공할 수 있다. 공유된 아이템은 채팅의 멤버들이 상호작용하여, 예를 들어, 대응하는 외부 리소스를 론칭하거나, 외부 리소스 내의 특정 정보를 보거나, 채팅의 멤버를 외부 리소스 내의 특정 위치 또는 상태로 데려갈 수 있는 상호작용 채팅 카드일 수 있다. 주어진 외부 리소스 내에서, 응답 메시지들은 메시징 클라이언트(104) 상의 사용자들에게 전송될 수 있다. 외부 리소스는 외부 리소스의 현재 컨텍스트에 기초하여 응답들에 상이한 미디어 아이템들을 선택적으로 포함할 수 있다.

[0030] 메시징 클라이언트(104)는 주어진 외부 리소스를 론칭하거나 액세스하기 위해 이용가능한 외부 리소스들(예를 들어, 제3자 또는 외부 애플리케이션들(109) 또는 애플릿들)의 리스트를 사용자에게 제시할 수 있다. 이 리스트는 컨텍스트 감지 메뉴(context-sensitive menu)에 제시될 수 있다. 예를 들어, 외부 애플리케이션(109)(또는 애플릿들) 중 상이한 것들을 나타내는 아이콘들은 메뉴가 (예를 들어, 대화 인터페이스로부터 또는 비-대화 인터페이스로부터) 사용자에 의해 어떻게 론칭되는지에 기초하여 달라질 수 있다.

[0031] 메시징 클라이언트(104)는 사용자로부터 수신된 음성 또는 음성 입력에 기초하여 하나 이상의 AR 경험을 사용자에게 제시할 수 있다. 예로서, 메시징 클라이언트(104)는 클라이언트 디바이스(102)에 의해 캡처된 이미지 또는 비디오에서 사람 또는 사용자를 검출할 수 있다. 메시징 클라이언트(104)는 사람에 대한 골격을 생성하여 신체의 골격 키포인트들을 식별할 수 있다. 골격 키포인트들에 기초하여, 메시징 클라이언트(104)는 이미지 또는 비디오에 묘사된 사람의 자세를 결정할 수 있다. 결정된 자세에 기초하여, 메시징 클라이언트(104)는 AR 동작을 수행하라는 커맨드을 포함하는 수신된 음성 입력의 의도를 명확히 하기 위해 하나 이상의 대명사 또는 부사를 식별할 수 있다. 이것은 사용자가 간단하고 모호한 음성 커맨드들을 입력할 수 있게 하고 메시징 클라이언트(104)는 사용자의 신체 자세를 포함하는 추가 정보에 기초하여 요청된 커맨드들을 정확하게 실행할 수 있다. 예로서, 메시징 클라이언트(104)는 "이것을 청색으로 변경"하는 것을 지정하는 음성 커맨드를 수신할 수 있다. 이 경우, 메시징 클라이언트(104)는 음성 입력이 대명사 "이것"이 지칭하는 객체가 모호하다고 결정할

수 있다. 메시징 클라이언트(104)는 음성 입력의 화자에 관한 자세 정보에 액세스하고 화자의 자세가 인물이 착용된 모자에 대한 참조를 포함한다고 결정할 수 있다. 예로서, 자세는 특정 방향을 가리키는 손을 포함한다. 메시징 클라이언트(104)는 특정 방향이 인물이 착용한 모자 또는 다른 패션 아이템과 교차한다고 결정할 수 있다. 이에 응답하여, 메시징 클라이언트(104)는 모호한 의도, 즉 대명사 "이것"을 화자가 가리키는 객체를, 이 경우에는, 모자 또는 다른 패션 아이템으로 해석한다. 그 다음, 메시징 클라이언트(104)는 모자 또는 다른 패션 아이템의 색상을 청색으로 변경하도록 요청된 AR 동작을 실행하거나 수행한다. 특히, 메시징 클라이언트(104)는 적합한 물품 또는 의류 세그멘테이션 프로세스를 사용하여 의류 물품을 세그멘테이션하고, 세그멘테이션의 내부에 있는 의류 물품의 픽셀들을 청색으로 변색시킨다. 음성 입력을 사용하는 AR 제어 프로세스의 추가 세부사항들은 도 5와 관련하여 아래에 제공된다.

[0032] 시스템 아키텍처

[0033] 도 2는 일부 예들에 따른 메시징 시스템(100)에 관한 추가 세부사항들을 예시하는 블록도이다. 구체적으로, 메시징 시스템(100)은 메시징 클라이언트(104) 및 애플리케이션 서버들(114)을 포함하는 것으로 도시된다. 메시징 시스템(100)은 메시징 클라이언트(104)에 의해 클라이언트측에서 그리고 애플리케이션 서버들(114)에 의해 서버측에서 지원되는 다수의 서브시스템들을 구현한다. 이러한 서브시스템들은, 예를 들어, 단기적 타이머 시스템(202), 컬렉션 관리 시스템(204), 증강 시스템(208), 맵 시스템(210), 게임 시스템(212) 및 외부 리소스 시스템(220)을 포함한다.

[0034] 단기적 타이머 시스템(202)은 메시징 클라이언트(104) 및 메시징 서버(118)에 의한 콘텐츠에 대한 일시적 또는 시간 제한적 액세스를 시행하는 것을 담당한다. 단기적 타이머 시스템(202)은 메시지 또는 메시지들의 컬렉션(예를 들어, 스토리)과 연관된 지속기간 및 디스플레이 파라미터들에 기초하여, 메시징 클라이언트(104)를 통해 메시지들 및 연관된 콘텐츠에 대한 (예를 들어, 제시 및 디스플레이를 위한) 액세스를 선택적으로 가능하게 하는 다수의 타이머를 포함한다. 단기적 타이머 시스템(202)의 동작에 관한 추가 세부사항들이 아래에 제공된다.

[0035] 컬렉션 관리 시스템(204)은 미디어의 컬렉션들(예를 들어, 텍스트, 이미지 비디오, 및 오디오 데이터의 컬렉션들) 또는 세트들을 관리하는 것을 담당한다. 콘텐츠의 컬렉션(예를 들어, 이미지들, 비디오, 텍스트, 및 오디오를 포함하는 메시지들)은 "이벤트 갤러리" 또는 "이벤트 스토리"로 조직될 수 있다. 이러한 컬렉션은, 콘텐츠가 관련된 이벤트의 지속기간과 같은, 지정된 시간 기간 동안 이용가능하게 될 수 있다. 예를 들어, 음악 콘서트에 관련된 콘텐츠는 그 음악 콘서트의 지속기간 동안 "스토리"로서 이용가능하게 될 수 있다. 컬렉션 관리 시스템(204)은 또한 메시징 클라이언트(104)의 사용자 인터페이스에 특정 컬렉션의 존재의 통지를 제공하는 아이콘을 게시하는 것을 담당할 수 있다.

[0036] 컬렉션 관리 시스템(204)은 또한 컬렉션 관리자가 콘텐츠의 특정 컬렉션을 관리 및 큐레이션하는 것을 허용하는 큐레이션 인터페이스(206)를 포함한다. 예를 들어, 큐레이션 인터페이스(206)는 이벤트 조직자가 특정 이벤트에 관련된 콘텐츠의 컬렉션을 큐레이션(예를 들어, 부적절한 콘텐츠 또는 중복 메시지들을 삭제)하는 것을 가능하게 한다. 추가적으로, 컬렉션 관리 시스템(204)은 머신 비전(또는 이미지 인식 기술) 및 콘텐츠 규칙들을 이용하여 콘텐츠 컬렉션을 자동으로 큐레이션한다. 특정 예들에서, 사용자 생성 콘텐츠를 컬렉션에 포함시키는 것에 대한 보상이 사용자에게 지불될 수 있다. 이러한 경우들에서, 컬렉션 관리 시스템(204)은 이러한 사용자들에게 그들의 콘텐츠를 사용하는 것에 대해 자동적으로 지불하도록 동작한다.

[0037] 증강 시스템(208)은 사용자가 메시지와 연관된 미디어 콘텐츠를 증강(예를 들어, 주석을 달거나 또는 다른 방식으로 수정 또는 편집)할 수 있게 하는 다양한 기능들을 제공한다. 예를 들어, 증강 시스템(208)은 메시징 시스템(100)에 의해 처리되는 메시지들에 대한 미디어 오버레이들의 생성 및 게시에 관련된 기능들을 제공한다. 증강 시스템(208)은 클라이언트 디바이스(102)의 지리위치에 기초하여 메시징 클라이언트(104)에 미디어 오버레이 또는 증강(예를 들어, 이미지 필터)을 동작적으로 공급한다. 다른 예에서, 증강 시스템(208)은 클라이언트 디바이스(102)의 사용자의 소셜 네트워크 정보와 같은 다른 정보에 기초하여 메시징 클라이언트(104)에 미디어 오버레이를 동작적으로 공급한다. 미디어 오버레이는 오디오 및 시각적 콘텐츠 및 시각적 효과들을 포함할 수 있다. 오디오 및 시각적 콘텐츠의 예들은 사진들, 텍스트들, 로고들, 애니메이션들, 및 음향 효과들을 포함한다. 시각적 효과의 예는 컬러 오버레잉을 포함한다. 오디오 및 시각적 콘텐츠 또는 시각적 효과는 클라이언트 디바이스(102)에서 미디어 콘텐츠 아이템(예컨대, 포토)에 적용될 수 있다. 예를 들어, 미디어 오버레이는 클라이언트 디바이스(102)에 의해 촬영된 사진의 위에 오버레이될 수 있는 텍스트, 그래픽 요소, 또는 이미지를 포함할 수 있다. 다른 예에서, 미디어 오버레이는 위치 식별 오버레이(예를 들어, 베니스 비치), 라이브 이벤트의 이름, 또는 상인 이름 오버레이(예를 들어, 비치 커피 하우스)를 포함한다. 다른 예에서, 증강 시스템(208)은

클라이언트 디바이스(102)의 지리위치를 사용하여 클라이언트 디바이스(102)의 지리위치에서 상인의 이름을 포함하는 미디어 오버레이를 식별한다. 미디어 오버레이는 상인과 연관된 다른 표시를 포함할 수 있다. 미디어 오버레이들은 데이터베이스(126)에 저장되고 데이터베이스 서버(120)를 통해 액세스될 수 있다.

[0038] 일부 예들에서, 증강 시스템(208)은 사용자들이 맵 상의 지리위치를 선택하고, 선택된 지리위치와 연관된 콘텐츠를 업로드할 수 있게 하는 사용자 기반 게시 플랫폼을 제공한다. 사용자는 또한, 특정 미디어 오버레이가 다른 사용자들에게 제공되어야 하는 상황을 지정할 수 있다. 증강 시스템(208)은, 업로드된 콘텐츠를 포함하는 그리고 업로드된 콘텐츠를 선택된 지리위치와 연관시키는 미디어 오버레이를 생성한다.

[0039] 다른 예들에서, 증강 시스템(208)은, 상인들이 입찰 프로세스를 통해 지리위치와 연관된 특정 미디어 오버레이를 선택할 수 있게 하는 상인 기반 게시 플랫폼을 제공한다. 예를 들어, 증강 시스템(208)은 최고 입찰 상인의 미디어 오버레이를 미리 정의된 시간의 양 동안 대응하는 지리위치와 연관시킨다. 증강 시스템(208)은 증강 현실 경험들을 획득하기 위해 이미지 처리 서버(122)와 통신하고, 하나 이상의 사용자 인터페이스에서 이러한 경험들의 식별자들을 (예를 들어, 실시간 이미지 또는 비디오를 통한 아이콘들로서 또는 증강 현실 경험들의 제시된 식별자들에 전용인 인터페이스들에서의 썸네일들 또는 아이콘들로서) 제시한다. AR 경험이 선택되면, 하나 이상의 이미지, 비디오, 또는 AR 그래픽 요소가 검색되고 클라이언트 디바이스(102)에 의해 캡처된 이미지들 또는 비디오 위에 오버레이로서 제시된다. 일부 경우들에서, 카메라는 전방 시야로 전환되고(예를 들어, 클라이언트 디바이스(102)의 전면 카메라가 특정 AR 경험의 활성화에 응답하여 활성화됨), 클라이언트 디바이스(102)의 후면 카메라 대신에 클라이언트 디바이스(102)의 전면 카메라로부터의 이미지들이 클라이언트 디바이스(102) 상에 디스플레이되기 시작한다. 하나 이상의 이미지, 비디오, 또는 AR 그래픽 요소가 검색되고, 클라이언트 디바이스(102)의 전면 카메라에 의해 캡처되고 디스플레이되는 이미지들의 위에 오버레이로서 제시된다.

[0040] 다른 예들에서, 증강 시스템(208)은 다른 클라이언트 디바이스(102) 상의 다른 증강 시스템(208)과 그리고 네트워크(112)를 통해 서버와 통신하고 데이터를 교환할 수 있다. 교환되는 데이터는 공유 AR 세션을 식별하는 세션 식별자, 공유 AR 세션을 공통 원점, 공통 좌표 프레임 및 기능들(예를 들어, 기능들을 기동하기 위한 커맨드들)뿐만 아니라 다른 페이로드 데이터(예를 들어, 텍스트, 오디오, 비디오 또는 다른 멀티미디어 데이터)에 정렬하기 위해 사용되는 제1 클라이언트 디바이스(102)와 제2 클라이언트 디바이스(102) 사이의 변환(예를 들어, 복수의 클라이언트 디바이스들(102)은 제1 및 제2 디바이스들을 포함함)을 포함할 수 있다.

[0041] 증강 시스템(208)은 제2 클라이언트 디바이스(102)가 변환에 기초하여 AR 좌표계를 조정할 수 있도록 변환을 제2 클라이언트 디바이스(102)에 전송한다. 이러한 방식으로, 제1 및 제2 클라이언트 디바이스들(102)은 AR 세션에서 콘텐츠를 디스플레이하기 위해 그들의 좌표계들 및 프레임들을 동기화한다. 구체적으로, 증강 시스템(208)은 제1 클라이언트 디바이스(102)의 좌표계에서 제2 클라이언트 디바이스(102)의 원점을 계산한다. 그 다음, 증강 시스템(208)은 제2 클라이언트 디바이스(102)의 좌표계에서의 제2 클라이언트 디바이스(102)의 관점으로부터의 원점의 위치에 기초하여 제2 클라이언트 디바이스(102)의 좌표계에서의 오프셋을 결정할 수 있다. 이 오프셋은 제2 클라이언트 디바이스(102)가 제1 클라이언트 디바이스(102)로서 공통 좌표계 또는 프레임에 따라 AR 콘텐츠를 생성하도록 변환을 생성하기 위해 사용된다.

[0042] 증강 시스템(208)은 개별 또는 공유 AR 세션들을 확립하기 위해 클라이언트 디바이스(102)와 통신할 수 있다. 증강 시스템(208)은 또한 공유 AR 세션에서 클라이언트 디바이스들(102)에 대한 전자 그룹 통신 세션(예를 들어, 그룹 채팅, 인스턴트 메시징)을 확립하기 위해 메시징 서버(118)에 결합될 수 있다. 전자 그룹 통신 세션은 전자 그룹 통신 세션 및 공유 AR 세션에 대한 액세스를 획득하기 위해 클라이언트 디바이스들(102)에 의해 제공되는 세션 식별자와 연관될 수 있다. 일 예에서, 클라이언트 디바이스들(102)은 먼저 전자 그룹 통신 세션에 대한 액세스를 획득한 다음, 클라이언트 디바이스들(102)이 공유 AR 세션에 액세스하는 것을 허용하는 전자 그룹 통신 세션에서의 세션 식별자를 획득한다. 일부 예들에서, 클라이언트 디바이스들(102)은 애플리케이션 서버들(114) 내의 증강 시스템(208)과의 보조 또는 통신 없이 공유 AR 세션에 액세스할 수 있다.

[0043] 맵 시스템(210)은 다양한 지리적 위치 기능들을 제공하고, 메시징 클라이언트(104)에 의한 맵 기반 미디어 콘텐츠 및 메시지들의 제시를 지원한다. 예를 들어, 맵 시스템(210)은, 맵의 컨텍스트 내에서, 맵 상에 사용자 아이콘들 또는 아바타들(예를 들어, 프로필 데이터(316)에 저장됨)의 디스플레이를 가능하게 하여 사용자의 "친구들"의 현재 또는 과거 위치뿐만 아니라 이러한 친구들에 의해 생성된 미디어 콘텐츠(예를 들어, 사진들 및 비디오들을 포함하는 메시지들의 컬렉션들)를 표시한다. 예를 들어, 특정 지리적 위치로부터 메시징 시스템(100)에 사용자에 의해 포스팅된 메시지는 메시징 클라이언트(104)의 맵 인터페이스 상의 특정 사용자의 "친구들"에게 그 특정 위치에서의 맵의 컨텍스트 내에 디스플레이될 수 있다. 사용자는 또한 메시징 클라이언트(104)를 통해

메시징 시스템(100)의 다른 사용자들과 자신의 위치 및 상태 정보를 (예를 들어, 적절한 상태 아바타를 사용하여) 공유할 수 있으며, 이 위치 및 상태 정보는 선택된 사용자들에 대한 메시징 클라이언트(104)의 맵 인터페이스의 컨텍스트 내에 유사하게 디스플레이된다.

[0044] 게임 시스템(212)은 메시징 클라이언트(104)의 컨텍스트 내에서 다양한 게이밍 기능들을 제공한다. 메시징 클라이언트(104)는 메시징 클라이언트(104)의 컨텍스트 내에서 사용자에 의해 론칭되고 메시징 시스템(100)의 다른 사용자들과 플레이될 수 있는 이용가능한 게임들(예를 들어, 웹 기반 게임들 또는 웹 기반 애플리케이션들)의 리스트를 제공하는 게임 인터페이스를 제공한다. 메시징 시스템(100)은 또한 메시징 클라이언트(104)로부터 다른 사용자들에게 초대들을 발행함으로써, 특정 사용자가 특정 게임의 플레이에 참여하도록 다른 사용자들을 초대할 수 있게 한다. 메시징 클라이언트(104)는 또한 게임플레이의 컨텍스트 내에서 음성 및 텍스트 메시징(예를 들어, 채팅들) 둘 다를 지원하고, 게임들에 대한 리더보드를 제공하고, 또한 게임 내 보상들(예를 들어, 코인들 및 아이템들)의 제공을 지원한다.

[0045] 외부 리소스 시스템(220)은 메시징 클라이언트(104)가 외부 리소스들을 론칭하거나 액세스하기 위해 외부 앱(들) 서버들(110)과 통신하기 위한 인터페이스를 제공한다. 각각의 외부 리소스(앱들) 서버(110)는, 예를 들어, 마크업 언어(예를 들어, HTML5) 기반 애플리케이션 또는 외부 애플리케이션의 소규모 버전(예를 들어, 메시징 클라이언트(104) 외부에 있는 게임, 유틸리티, 지불, 또는 승차 공유 애플리케이션)을 호스팅한다. 메시징 클라이언트(104)는 웹 기반 리소스와 연관된 외부 리소스(앱들) 서버들(110)로부터 HTML5 파일에 액세스함으로써 웹 기반 리소스(예를 들어, 애플리케이션)를 론칭할 수 있다. 특정 예들에서, 외부 리소스 서버들(110)에 의해 호스팅되는 애플리케이션들은 메시징 서버(118)에 의해 제공되는 SDK(Software Development Kit)를 활용하여 자바스크립트로 프로그램된다. SDK는 웹 기반 애플리케이션에 의해 호출되거나 기동될 수 있는 기능들을 갖는 API들을 포함한다. 특정 예들에서, 메시징 서버(118)는 메시징 클라이언트(104)의 특정 사용자 데이터에 대한 주어진 제3자 리소스 액세스를 제공하는 자바스크립트 라이브러리를 포함한다. HTML5가 게임들을 프로그램하기 위한 예시적인 기술로서 사용되지만, 다른 기술들에 기초하여 프로그램되는 애플리케이션들 및 리소스들이 사용될 수 있다.

[0046] SDK의 기능들을 웹 기반 리소스에 통합하기 위해, SDK는 메시징 서버(118)로부터 외부 리소스(앱들) 서버(110)에 의해 다운로드되거나 또는 달리 외부 리소스(앱들) 서버(110)에 의해 수신된다. 일단 다운로드되거나 수신되면, SDK는 웹 기반 외부 리소스의 애플리케이션 코드의 일부로서 포함된다. 웹 기반 리소스의 코드는 이어서 메시징 클라이언트(104)의 특징들을 웹 기반 리소스에 통합하기 위해 SDK의 특정 기능들을 호출하거나 기동할 수 있다.

[0047] 메시징 서버(118) 상에 저장된 SDK는 외부 리소스(예를 들어, 제3자 또는 외부 애플리케이션들(109) 또는 애플릿들과 메시징 클라이언트(104)) 사이의 브리지를 효과적으로 제공한다. 이는 메시징 클라이언트(104) 상의 다른 사용자들과 통신하는 매끄러운 경험을 사용자에게 제공하는 한편, 또한 메시징 클라이언트(104)의 모습 및 느낌을 보존한다. 외부 리소스와 메시징 클라이언트(104) 사이의 통신들을 브리징하기 위해, 특정 예들에서, SDK는 외부 리소스 서버들(110)과 메시징 클라이언트(104) 사이의 통신을 용이하게 한다. 특정 예들에서, 클라이언트 디바이스(102) 상에서 실행되는 WebViewJavaScriptBridge는 외부 리소스와 메시징 클라이언트(104) 사이에 2개의 단방향 통신 채널을 확립한다. 메시지들은 이러한 통신 채널들을 통해 외부 리소스와 메시징 클라이언트(104) 사이에서 비동기적으로 전송된다. 각각의 SDK 기능 호출은 메시지 및 콜백(callback)으로서 전송된다. 각각의 SDK 기능은 고유 콜백 식별자를 구성하는 것 및 그 콜백 식별자를 갖는 메시지를 전송하는 것에 의해 구현된다.

[0048] SDK를 사용함으로써, 메시징 클라이언트(104)로부터의 모든 정보가 외부 리소스 서버들(110)과 공유되는 것은 아니다. SDK는 외부 리소스의 요구들에 기초하여 어느 정보가 공유되는지를 제한한다. 특정 예들에서, 각각의 외부 리소스 서버(110)는 웹 기반 외부 리소스에 대응하는 HTML5 파일을 메시징 서버(118)에 제공한다. 메시징 서버(118)는 메시징 클라이언트(104)에서 웹 기반 외부 리소스의 (박스 아트 또는 다른 그래픽과 같은) 시각적 표현을 추가할 수 있다. 일단 사용자가 시각적 표현을 선택하거나 메시징 클라이언트(104)의 GUI를 통해 메시징 클라이언트(104)에게 웹 기반 외부 리소스의 특징들에 액세스하도록 지시하면, 메시징 클라이언트(104)는 HTML5 파일을 획득하고 웹 기반 외부 리소스의 특징들에 액세스하는 데 필요한 리소스들을 인스턴스화한다.

[0049] 메시징 클라이언트(104)는 외부 리소스에 대한 그래픽 사용자 인터페이스(예를 들어, 랜딩 페이지 또는 타이틀 스크린)를 제시한다. 랜딩 페이지 또는 타이틀 스크린을 제시하는 동안, 그 전에, 또는 그 후에, 메시징 클라이언트(104)는 론칭된 외부 리소스가 메시징 클라이언트(104)의 사용자 데이터에 액세스하도록 이전에 인가되었

는지를 결정한다. 론칭된 외부 리소스가 메시징 클라이언트(104)의 사용자 데이터에 액세스하도록 이전에 인가되었다고 결정하는 것에 응답하여, 메시징 클라이언트(104)는 외부 리소스의 기능들 및 특징들을 포함하는 외부 리소스의 다른 그래픽 사용자 인터페이스를 제시한다. 론칭된 외부 리소스가 메시징 클라이언트(104)의 사용자 데이터에 액세스하도록 이전에 인가되지 않았다고 결정하는 것에 응답하여, 외부 리소스의 랜딩 페이지 또는 타이틀 스크린을 디스플레이하는 임계 시간 기간(예를 들어, 3초) 후에, 메시징 클라이언트(104)는 사용자 데이터에 액세스하도록 외부 리소스를 인가하기 위한 메뉴를 위로 슬라이딩한다(예를 들어, 메뉴를 스크린의 하단으로부터 스크린의 중간 또는 다른 부분으로 표면화하는 것으로서 애니메이션화한다). 메뉴는 외부 리소스가 사용하도록 인가될 사용자 데이터의 타입을 식별해 준다. 수용 옵션의 사용자 선택을 수신하는 것에 응답하여, 메시징 클라이언트(104)는 외부 리소스를 인가된 외부 리소스들의 리스트에 추가하고 외부 리소스가 메시징 클라이언트(104)로부터의 사용자 데이터에 액세스하는 것을 허용한다. 일부 예들에서, 외부 리소스는 OAuth 2 프레임워크에 따라 사용자 데이터에 액세스하도록 메시징 클라이언트(104)에 의해 인가된다.

[0050] 메시징 클라이언트(104)는 인가되는 외부 리소스의 타입에 기초하여 외부 리소스들과 공유되는 사용자 데이터의 타입을 제어한다. 예를 들어, 풀-스케일 외부 애플리케이션들(예를 들어, 제3자 또는 외부 애플리케이션(109))을 포함하는 외부 리소스들에는 제1 타입의 사용자 데이터(예를 들어, 상이한 아바타 특성들을 갖거나 갖지 않는 사용자들의 2D 아바타들만)에 대한 액세스가 제공된다. 또 다른 예로서, 외부 애플리케이션들의 소규모 버전들(예를 들어, 제3자 애플리케이션들의 웹 기반 버전들)을 포함하는 외부 리소스들에게는 제2 타입의 사용자 데이터(예를 들어, 지불 정보, 사용자들의 2D 아바타들, 사용자들의 3D 아바타들, 및 다양한 아바타 특성들을 갖는 아바타들)에 대한 액세스가 제공된다. 아바타 특성들은 상이한 자세들, 얼굴 특징들, 의류 등과 같은 아바타의 모습 및 느낌을 맞춤화하는 상이한 방식들을 포함한다.

[0051] 의도 결정 시스템(224)은 하나 이상의 사용자를 묘사하는 이미지 및 음성 입력을 수신한다. 의도 결정 시스템(224)은 음성 입력이 모호한 의도를 포함한다고(예를 들어, 대명사 또는 부사를 포함한다고) 결정한다. 의도 결정 시스템(224)은 모호한 의도를 해석하거나 분별하기 위해 이미지에 묘사된 하나 이상의 사용자의 자세를 결정한다. 모호한 의도가 해석되거나 식별되면, 의도 결정 시스템(224)은 음성 입력의 모호한 의도가 참조하는 객체, 대상, 아이템, 또는 방향에 기초하여 음성 입력에서 요청된 하나 이상의 AR 동작을 수행한다. 일 예에서, 의도 결정 시스템(224)은 하나 이상의 AR 요소를 사용하여 적어도 하나의 현실 세계 객체를 변경, 혼합, 변형, 변환, 또는 모핑한다.

[0052] 구체적으로, 의도 결정 시스템(224)은 AR/VR 동작을 수행하라는 구두 커맨드에 응답하여 클라이언트 디바이스(102) 상에 구현되는 AR/VR 애플리케이션에 의해 액세스될 수 있는 컴포넌트이다. AR/VR 애플리케이션은 RGB 카메라를 사용하여 사용자의 단안 이미지를 캡처한다. AR/VR 애플리케이션은 사용자의 캡처된 이미지에 다양한 훈련된 머신 러닝 기술들을 적용하여 이미지에 묘사된 사용자의 자세를 결정하여 하나 이상의 구두 커맨드를 명확하게 하고 구두 커맨드들에 기초하여 캡처된 이미지에 하나 이상의 AR 시각적 효과를 적용한다.

[0053] 훈련에서, 의도 결정 시스템(224)은 모호한 의도(예를 들어, 대명사 또는 부사)를 갖는 음성 입력 및 상이한 자세들을 갖는 하나 이상의 사용자의 묘사들을 포함하는 제1 복수의 입력 훈련 이미지들을 획득한다. 이러한 훈련 이미지들은 또한 음성 입력의 모호한 의도에 대응하는 객체, 방향, 아이템, 또는 대상을 포함하는 실측 정보를 제공한다. 머신 러닝 기술(예를 들어, 심층 신경망)은 복수의 훈련 이미지들의 특징들에 기초하여 훈련된다. 구체적으로, 머신 러닝 기술은 주어진 훈련 이미지로부터 하나 이상의 특징을 추출하고, 주어진 훈련 이미지에 묘사된 자세에 대한 음성 입력의 모호한 의도에 대응하는 객체, 방향, 아이템, 또는 대상을 추정한다. 머신 러닝 기술은 훈련 이미지에 대응하는 객체, 방향, 아이템, 또는 대상을 포함하는 실측 정보를 획득하고, 의도(예를 들어, 음성 입력의 모호한 의도에 대응하는 객체, 방향, 아이템, 또는 대상)의 후속 추정들을 개선하기 위해 하나 이상의 계수 또는 파라미터를 조정 또는 업데이트한다. 머신 러닝 기술은 하나 이상의 중지 기준이 충족될 때까지 추가적인 훈련 데이터를 처리하고 파라미터들을 업데이트하는 것을 계속한다.

[0054] 데이터 아키텍처

[0055] 도 3은 특정 예들에 따른, 메시징 서버 시스템(108)의 데이터베이스(126)에 저장될 수 있는 데이터 구조들(300)을 예시하는 개략도이다. 데이터베이스(126)의 콘텐츠가 다수의 테이블을 포함하는 것으로 도시되어 있지만, 데이터가 다른 타입들의 데이터 구조들로 (예를 들어, 객체 지향 데이터베이스로서) 저장될 수 있다는 것이 이해될 것이다.

[0056] 데이터베이스(126)는 메시지 테이블(302) 내에 저장된 메시지 데이터를 포함한다. 이 메시지 데이터는, 임의의 특정한 하나의 메시지에 대해, 적어도 메시지 발신자 데이터, 메시지 수신인(또는 수신자) 데이터, 및 페이로드

를 포함한다. 메시지에 포함될 수 있고 메시지 테이블(302)에 저장된 메시지 데이터 내에 포함될 수 있는 정보에 관한 추가 세부사항들은 도 4를 참조하여 아래에 설명된다.

[0057] 엔티티 테이블(306)은 엔티티 데이터를 저장하고, 엔티티 그래프(308) 및 프로필 데이터(316)에 (예를 들어, 참조식으로) 링크된다. 그에 대해 레코드들이 엔티티 테이블(306) 내에 유지되는 엔티티들은 개인, 법인 엔티티, 조직, 객체, 장소, 이벤트 등을 포함할 수 있다. 엔티티 타입에 관계없이, 메시징 서버 시스템(108)이 데이터를 저장하는 것에 관한 임의의 엔티티는 인식된 엔티티일 수 있다. 각각의 엔티티에는 고유 식별자는 물론, 엔티티 타입 식별자(도시되지 않음)가 제공된다.

[0058] 엔티티 그래프(308)는 엔티티들 사이의 관계 및 연관에 관한 정보를 저장한다. 이러한 관계들은, 단지 예를 들어, 사회적, 전문적(예를 들어, 일반 법인 또는 조직에서의 일) 관심 기반 또는 활동 기반일 수 있다.

[0059] 프로필 데이터(316)는 특정 엔티티에 대한 다수의 타입의 프로필 데이터를 저장한다. 프로필 데이터(316)는 특정 엔티티에 의해 지정되는 프라이버시 설정들에 기초하여, 메시징 시스템(100)의 다른 사용자들에게 선택적으로 사용 및 제시될 수 있다. 엔티티가 개인인 경우, 프로필 데이터(316)는, 예를 들어, 사용자 이름, 전화 번호, 주소 및 설정들(예를 들어, 통지 및 프라이버시 설정들)은 물론, 사용자-선택된 아바타 표현(또는 이러한 아바타 표현들의 컬렉션)을 포함한다. 그 다음, 특정 사용자는 메시징 시스템(100)을 통해 통신되는 메시지들의 콘텐츠 내에, 그리고 메시징 클라이언트들(104)에 의해 다른 사용자들에게 디스플레이되는 맵 인터페이스들 상에 이러한 아바타 표현들 중 하나 이상을 선택적으로 포함할 수 있다. 아바타 표현들의 컬렉션은 사용자가 특정 시간에 통신하기 위해 선택할 수 있는 상태 또는 활동의 그래픽 표현을 제시하는 "상태 아바타들"을 포함할 수 있다.

[0060] 엔티티가 그룹인 경우, 그룹에 대한 프로필 데이터(316)는, 관련 그룹에 대한 그룹 이름, 멤버들, 및 다양한 설정들(예를 들어, 통지들) 외에도, 그룹과 연관된 하나 이상의 아바타 표현을 유사하게 포함할 수 있다.

[0061] 데이터베이스(126)는 또한 오버레이들 또는 필터들과 같은 증강 데이터를 증강 테이블(310)에 저장한다. 증강 데이터는 비디오들(데이터가 비디오 테이블(304)에 저장됨) 및 이미지들(데이터가 이미지 테이블(312)에 저장됨)과 연관되고 이들에 적용된다.

[0062] 데이터베이스(126)는 또한 개별 및 공유 AR 세션들에 관련한 데이터를 저장할 수 있다. 이 데이터는 제1 클라이언트 디바이스(102)의 AR 세션 클라이언트 제어기와 제2 클라이언트 디바이스(102)의 다른 AR 세션 클라이언트 제어기 사이에서 통신되는 데이터, 및 AR 세션 클라이언트 제어기와 증강 시스템(208) 사이에서 통신되는 데이터를 포함할 수 있다. 데이터는 공유된 AR 장면의 공통 좌표 프레임, 디바이스들 사이의 변환, 세션 식별자, 신체, 골격 관절 위치들, 손목 관절 위치들, 발 등을 묘사하는 이미지들을 확립하기 위해 사용되는 데이터를 포함할 수 있다.

[0063] 일 예에서, 필터들은, 수신 사용자로의 제시 동안 이미지 또는 비디오상에 오버레이되어 디스플레이되는 오버레이들이다. 필터들은, 전송 사용자가 메시지를 작성하고 있을 때 메시징 클라이언트(104)에 의해 전송 사용자에게 제시되는 필터들의 세트로부터의 사용자-선택된 필터들을 비롯하여, 다양한 타입들일 수 있다. 다른 타입들의 필터들은 지리적 위치에 기초하여 전송 사용자에게 제시될 수 있는 지리위치 필터들(지오-필터들이라고도 알려짐)을 포함한다. 예를 들어, 이웃 또는 특수 위치에 특정한 지리위치 필터들이 클라이언트 디바이스(102)의 GPS(Global Positioning System) 유닛에 의해 결정된 지리위치 정보에 기초하여 메시징 클라이언트(104)에 의해 사용자 인터페이스 내에 제시될 수 있다.

[0064] 다른 타입의 필터는 메시지 생성 프로세스 동안 클라이언트 디바이스(102)에 의해 수집된 다른 입력들 또는 정보에 기초하여, 메시징 클라이언트(104)에 의해 전송 사용자에게 선택적으로 제시될 수 있는 데이터 필터이다. 데이터 필터들의 예들은 특정 위치에서의 현재 온도, 전송 사용자가 이동하고 있는 현재 속도, 클라이언트 디바이스(102)에 대한 배터리 수명, 또는 현재 시간을 포함한다.

[0065] 이미지 테이블(312) 내에 저장될 수 있는 다른 증강 데이터는 (예를 들어, 증강 현실 경험들을 적용하는 것에 대응하는) AR 콘텐츠 아이템들을 포함한다. AR 콘텐츠 아이템 또는 AR 아이템은 이미지 또는 비디오에 추가될 수 있는 실시간 특수 효과 및 사운드일 수 있다.

[0066] 위에서 설명된 바와 같이, 증강 데이터는 AR 콘텐츠 아이템들, 오버레이들, 이미지 변환들, AR 이미지들, AR 로고 또는 엠블럼들 및 이미지 데이터(예를 들어, 비디오들 또는 이미지들)에 적용될 수 있는 수정들을 지칭하는 유사한 용어들을 포함한다. 이는 클라이언트 디바이스(102)의 디바이스 센서들(예를 들어, 하나 또는 다수의 카메라)을 사용하여 캡처될 때 이미지를 수정한 다음 수정들과 함께 클라이언트 디바이스(102)의 스크린 상에

디스플레이되는 실시간 수정들을 포함한다. 이는 또한 수정될 수 있는 갤러리 내의 비디오 클립들과 같은 저장된 콘텐츠에 대한 수정들을 포함한다. 예를 들어, 다수의 AR 콘텐츠 아이템들에 대한 액세스를 갖는 클라이언트 디바이스(102)에서, 사용자는 다수의 AR 콘텐츠 아이템들을 갖는 단일 비디오 클립을 사용하여 상이한 AR 콘텐츠 아이템들이 저장된 클립을 어떻게 수정할지를 볼 수 있다. 예를 들어, 콘텐츠에 대한 상이한 AR 콘텐츠 아이템들을 선택함으로써 동일한 콘텐츠에 상이한 의사 랜덤 움직임 모델들을 적용하는 다수의 AR 콘텐츠 아이템들이 적용될 수 있다. 유사하게, 실시간 비디오 캡처는 클라이언트 디바이스(102)의 센서들에 의해 현재 캡처되고 있는 비디오 이미지들이 캡처된 데이터를 어떻게 수정할지를 보여주기 위해 예시된 수정과 함께 사용될 수 있다. 이러한 데이터는 단순히 스크린 상에 디스플레이되고 메모리에 저장되지 않을 수 있거나, 또는 디바이스 센서들에 의해 캡처된 콘텐츠가 수정들과 함께 또는 수정들 없이(또는 둘 다) 메모리에 기록되고 저장될 수 있다. 일부 시스템들에서, 미리보기 특징은 상이한 AR 콘텐츠 아이템들이 동시에 디스플레이 내의 상이한 윈도우들 내에서 어떤 모습으로 보일지를 보여줄 수 있다. 이는, 예를 들어, 상이한 의사 랜덤 애니메이션들을 갖는 다수의 윈도우가 디스플레이 상에서 동시에 보여지는 것을 가능하게 할 수 있다.

[0067] 따라서, 데이터 및 이 데이터를 사용하여 콘텐츠를 수정하는 AR 콘텐츠 아이템들을 사용하는 다양한 시스템들 또는 다른 이러한 변환 시스템들은 현실 세계 객체들(예를 들어, 얼굴들, 손들, 신체들, 고양이들, 개들, 표면들, 객체들 등)의 검출, 이러한 객체들이 비디오 프레임들에서 시야를 떠나고, 진입하고, 그 주위를 이동할 때 이러한 객체들의 추적, 및 이러한 객체들이 추적될 때 이러한 객체들의 수정 또는 변환을 수반할 수 있다. 다양한 예들에서, 이러한 변환들을 달성하기 위한 상이한 방법들이 사용될 수 있다. 일부 예들은 객체 또는 객체들의 3D 메시 모델을 생성하는 것, 및 변환을 달성하기 위해 비디오 내의 모델의 변환들 및 애니메이션화된 텍스처들을 사용하는 것을 수반할 수 있다. 다른 예들에서, 객체 상의 포인트들의 추적은 추적된 위치에 이미지 또는 텍스처(이는 2D 또는 3D일 수 있음)를 배치하기 위해 이용될 수 있다. 또 다른 예들에서, 비디오 프레임들의 신경망 분석은 콘텐츠(예를 들어, 비디오의 이미지들 또는 프레임들)에 이미지들, 모델들, 또는 텍스처들을 배치하기 위해 사용될 수 있다. 따라서, AR 콘텐츠 아이템들 또는 요소들은 콘텐츠에서 변환들을 생성하기 위해 사용되는 이미지들, 모델들, 및 텍스처들뿐만 아니라, 객체 검출, 추적, 및 배치로 이러한 변환들을 달성하기 위해 필요한 추가적인 모델링 및 분석 정보 둘 다를 지칭한다.

[0068] 실시간 비디오 처리는 임의의 종류의 컴퓨터화된 시스템의 메모리에 저장된 임의의 종류의 비디오 데이터(예를 들어, 비디오 스트림들, 비디오 파일들 등)로 수행될 수 있다. 예를 들어, 사용자는 비디오 파일들을 로딩하고 이들을 디바이스의 메모리에 저장할 수 있거나, 또는 디바이스의 센서들을 사용하여 비디오 스트림을 생성할 수 있다. 추가적으로, 임의의 객체들은 인간의 얼굴 및 인체의 부분들, 동물들, 또는 의자들, 자동차들, 또는 다른 객체들과 같은 무생물들과 같은 컴퓨터 애니메이션 모델을 이용하여 처리될 수 있다.

[0069] 일부 예들에서, 변환될 콘텐츠와 함께 특정 수정이 선택될 때, 변환될 요소들이 컴퓨팅 디바이스에 의해 식별되고, 그 후 이들이 비디오의 프레임들에 존재하는 경우 검출되고 추적된다. 객체의 요소들은 수정 요청에 따라 수정되고, 따라서 비디오 스트림의 프레임들을 변환한다. 비디오 스트림의 프레임들의 변환은 상이한 종류들의 변환을 위한 상이한 방법들에 의해 수행될 수 있다. 예를 들어, 객체의 요소들의 형태들을 변경하는 것을 주로 참조하는 프레임들의 변환들에 대해, 객체의 각각의 요소에 대한 특성 포인트들이 (예를 들어, ASM(Active Shape Model) 또는 다른 공지된 방법들을 사용하여) 계산된다. 그 후, 객체의 적어도 하나의 요소 각각에 대해 특징적인 포인트들에 기초한 메시가 생성된다. 이 메시는 비디오 스트림에서 객체의 요소들을 추적하는 다음 스테이지에서 사용된다. 추적 프로세스에서, 각각의 요소에 대한 언급된 메시는 각각의 요소의 위치와 정렬된다. 그 후, 추가 포인트들이 메시 상에 생성된다. 수정 요청에 기초하여 각각의 요소에 대해 제1 포인트의 제1 세트가 생성되고, 제1 포인트 세트 및 수정 요청에 기초하여 각각의 요소에 대해 제2 포인트 세트가 생성된다. 그 후, 제1 및 제2 포인트들의 세트들 및 메시에 기초하여 객체의 요소들을 수정함으로써 비디오 스트림의 프레임들이 변환될 수 있다. 이러한 방법에서, 배경을 추적하고 수정함으로써 수정된 객체의 배경이 또한 변경되거나 왜곡될 수 있다.

[0070] 일부 예들에서, 객체의 요소들을 이용하여 객체의 일부 영역들을 변경하는 변환들은 객체의 각각의 요소에 대한 특징적인 포인트들을 계산하고, 계산된 특징적인 포인트들에 기초하여 메시를 생성함으로써 수행될 수 있다. 메시 상에 포인트들이 생성되고, 그 후 포인트들에 기초한 다양한 영역들이 생성된다. 그 다음, 각각의 요소에 대한 영역을 적어도 하나의 요소 각각에 대한 위치와 정렬함으로써 객체의 요소들이 추적되고, 영역들의 특성들은 수정을 위한 요청에 기초하여 수정될 수 있고, 따라서 비디오 스트림의 프레임들을 변환한다. 수정에 대한 특정 요청에 따라, 언급된 영역들의 특성들이 상이한 방식들로 변환될 수 있다. 이러한 수정들은 영역들의 색상을 변경하는 것; 비디오 스트림의 프레임들로부터 영역들의 적어도 일부 부분을 제거하는 것; 수정 요청에 기

초하는 영역들에 하나 이상의 새로운 객체를 포함시키는 것; 및 영역 또는 객체의 요소들을 수정하거나 왜곡하는 것을 수반할 수 있다. 다양한 예들에서, 이러한 수정들 또는 다른 유사한 수정들의 임의의 조합이 사용될 수 있다. 애니메이션화될 특정 모델들에 대해, 일부 특징적인 포인트들이 모델 애니메이션에 대한 옵션들의 전체 상태 공간을 결정하는데 이용될 제어 포인트들로서 선택될 수 있다.

[0071] 신체/사람 검출을 사용하여 이미지 데이터를 변환하기 위한 컴퓨터 애니메이션 모델의 일부 예들에서, 신체/사람은 특정 신체/사람 검출 알고리즘(예를 들어, 3D 인간 자세 추정 및 메시 재구성 프로세스들)을 사용하여 이미지 상에서 검출된다. 그 후, 이미지의 신체/사람 영역에 ASM 알고리즘이 적용되어 신체/사람 특징 기준 포인트들을 검출한다.

[0072] 신체/사람 검출에 적절한 다른 방법들 및 알고리즘들이 사용될 수 있다. 예를 들어, 일부 예들에서, 특징들은 고려 중인 이미지들의 대부분에 존재하는 구별가능한 포인트를 나타내는 랜드마크를 사용하여 위치된다. 신체/사람 랜드마크들에 대해, 예를 들어, 좌측 팔의 위치가 사용될 수 있다. 초기 랜드마크가 식별가능하지 않은 경우, 2차 랜드마크들이 사용될 수 있다. 이러한 랜드마크 식별 절차들은 임의의 이러한 객체들에 대해 사용될 수 있다. 일부 예들에서, 랜드마크들의 세트는 형상을 형성한다. 형상들은 형상 내의 포인트들의 좌표들을 사용하여 벡터들로서 표현될 수 있다. 어느 한 형상은 형상 포인트들 사이의 평균 유클리드 거리를 최소화하는 유사성 변환(similarity transform)(병진, 스케일링 및 회전을 허용함)을 이용하여 또 다른 형상에 정렬된다. 평균(mean) 형상은 정렬된 훈련 형상들의 평균이다.

[0073] 일부 예들에서, 전역적 신체/사람 검출기에 의해 결정된 신체/사람의 위치 및 크기에 정렬된 평균 형상으로부터 랜드마크들에 대한 검색이 시작된다. 그 후 이러한 검색은 각각의 포인트 주변의 이미지 텍스처의 템플릿 매칭에 의해 형상 포인트들의 위치들을 조정함으로써 임시 형상을 제안하는 단계 및 그 후 수렴이 발생할 때까지 임시 형상을 전역적 형상 모델에 일치시키는 단계를 반복한다. 일부 시스템들에서, 개별 템플릿 매칭들은 신뢰할 수 없고, 형상 모델은 약한 템플릿 매칭들의 결과들을 풀링하여 더 강한 전체 분류기를 형성한다. 전체 검색은 조대 해상도로부터 미세 해상도까지 이미지 피라미드에서의 각각의 레벨에서 반복된다.

[0074] 변환 시스템은 클라이언트 디바이스(예를 들어, 클라이언트 디바이스(102)) 상에서 이미지 또는 비디오 스트림을 캡처하고, 적합한 사용자 경험, 계산 시간, 및 전력 소비를 유지하면서 클라이언트 디바이스(102) 상에서 로컬로 복잡한 이미지 조작들을 수행할 수 있다. 복잡한 이미지 조작들은, 크기 및 형상 변화들, 감정 전이(예를 들어, 얼굴을 찡그림에서 미소로 변화시킴), 상태 전이(예를 들어, 대상의 노화, 외견상 나이 감소, 성별 변화), 스타일 전이, 그래픽 요소 애플리케이션, 3D 인간 자세 추정, 3D 신체 메시 재구성 및 클라이언트 디바이스(102) 상에서 효율적으로 실행되도록 구성되는 컨볼루션 신경망에 의해 구현된 임의의 다른 적합한 이미지 또는 비디오 조작을 포함할 수 있다.

[0075] 일부 예들에서, 이미지 데이터를 변환하기 위한 컴퓨터 애니메이션 모델은 사용자가 클라이언트 디바이스(102) 상에서 동작하는 메시징 클라이언트(104)의 일부로서 동작하는 신경망을 갖는 클라이언트 디바이스(102)를 사용하여 사용자의 이미지 또는 비디오 스트림(예를 들어, 셀카)을 캡처할 수 있는 시스템에 의해 사용될 수 있다. 메시징 클라이언트(104) 내에서 동작하는 변환 시스템이 이미지 또는 비디오 스트림 내의 신체/사람의 존재를 결정하고 이미지 데이터를 변환하기 위해 컴퓨터 애니메이션 모델과 연관된 수정 아이콘들을 제공하거나, 컴퓨터 애니메이션 모델은 본 명세서에 설명된 인터페이스와 연관된 것으로서 존재할 수 있다. 수정 아이콘들은 수정 동작의 일부로서 이미지 또는 비디오 스트림 내의 사용자의 신체/사람을 수정하기 위한 기초일 수 있는 변경들을 포함한다. 수정 아이콘이 선택되면, 변환 시스템은 선택된 수정 아이콘을 반영하기 위해 사용자의 이미지를 변환하는(예를 들어, 사용자 상에 미소짓는 얼굴을 생성하는) 프로세스를 개시한다. 수정된 이미지 또는 비디오 스트림은 이미지 또는 비디오 스트림이 캡처되자마자 클라이언트 디바이스(102) 상에 디스플레이되는 그래픽 사용자 인터페이스에 제시될 수 있고, 지정된 수정이 선택된다. 변환 시스템은 선택된 수정을 생성하고 적용하기 위해 이미지 또는 비디오 스트림의 일부에 대해 복잡한 컨볼루션 신경망을 구현할 수 있다. 즉, 사용자는 이미지 또는 비디오 스트림을 캡처하고 수정 아이콘이 선택되면 실시간으로 또는 거의 실시간으로 수정된 결과를 제시받을 수 있다. 또한, 수정은 비디오 스트림이 캡처되고 있는 동안 지속적일 수 있고, 선택된 수정 아이콘은 토글링된 채로 유지된다. 머신-교시 신경망(machine-taught neural network)들이 이러한 수정들을 가능하게 하기 위해 사용될 수 있다.

[0076] 변환 시스템에 의해 수행되는 수정을 제시하는 그래픽 사용자 인터페이스는 사용자에게 추가 상호작용 옵션들을 공급할 수 있다. 이러한 옵션들은 특정 컴퓨터 애니메이션 모델의 콘텐츠 캡처 및 선택을 개시(예를 들어, 콘텐츠 생성자 사용자 인터페이스로부터 개시)하기 위해 사용되는 인터페이스에 기초할 수 있다. 다양한 예들에

서, 수정은 수정 아이콘의 초기 선택 후에 지속적일 수 있다. 사용자는 변환 시스템에 의해 수정되고 있는 신체/사람을 탭핑하거나 다른 방식으로 선택함으로써 수정을 온 또는 오프로 토글링하고, 나중에 보기 위해 이것을 저장하거나 또는 이미징 애플리케이션의 다른 영역들로 브라우징할 수 있다. 다수의 얼굴들이 변환 시스템에 의해 수정되는 경우, 사용자는 그래픽 사용자 인터페이스 내에서 수정되고 디스플레이되는 단일 신체/사람을 탭핑하거나 선택함으로써 전역적으로 수정을 온 또는 오프로 토글링할 수 있다. 일부 예들에서, 다중의 신체/사람의 그룹 중 개별 신체/사람들은 개별적으로 수정될 수 있거나, 또는 이러한 수정들은 그래픽 사용자 인터페이스 내에 디스플레이된 개별 신체/사람 또는 일련의 개별 신체/사람들을 탭핑하거나 선택함으로써 개별적으로 토글링될 수 있다.

[0077] 스토리 테이블(314)은 컬렉션(예를 들어, 스토리 또는 갤러리)으로 컴파일되는 메시지들 및 연관된 이미지, 비디오, 또는 오디오 데이터의 컬렉션들에 관한 데이터를 저장한다. 특정 컬렉션의 생성은 특정 사용자(예를 들어, 레코드가 엔티티 테이블(306)에 유지되는 각각의 사용자)에 의해 개시될 수 있다. 사용자는 그 사용자에 의해 생성되고 전송/브로드캐스트된 콘텐츠의 컬렉션의 형태로 "개인 스토리"를 생성할 수 있다. 이를 위해, 메시징 클라이언트(104)의 사용자 인터페이스는 전송 사용자가 자신의 개인 스토리에 특정 콘텐츠를 추가할 수 있게 하기 위해 사용자 선택가능한 아이콘을 포함할 수 있다.

[0078] 컬렉션은 또한 수동으로, 자동으로, 또는 수동 및 자동 기술들의 조합을 사용하여 생성되는 다수의 사용자들로부터의 콘텐츠의 컬렉션인 "라이브 스토리"를 구성할 수 있다. 예를 들어, "라이브 스토리"는 다양한 위치들 및 이벤트들로부터의 사용자-제출 콘텐츠의 큐레이션된 스트림을 구성할 수 있다. 클라이언트 디바이스들이 위치 서비스 가능하고 특정 시간에 공통 위치 이벤트에 있는 사용자들에게는, 예를 들어, 메시징 클라이언트(104)의 사용자 인터페이스를 통해, 특정 라이브 스토리에 콘텐츠를 기여하는 옵션이 제시될 수 있다. 라이브 스토리는 그 또는 그녀의 위치에 기초하여 메시징 클라이언트(104)에 의해 사용자에게 식별될 수 있다. 최종 결과는 커뮤니티 관점에서 말하는 "라이브 스토리"이다.

[0079] 추가적인 타입의 콘텐츠 컬렉션은 "위치 스토리(location story)"로서 알려져 있으며, 이는 클라이언트 디바이스(102)가 특정 지리적 위치 내에(예를 들어, 대학 또는 대학 캠퍼스 상에) 위치되는 사용자가 특정 컬렉션에 기여할 수 있게 한다. 일부 예들에서, 위치 스토리에 대한 기여는 최종 사용자가 특정 조직 또는 다른 엔티티에 속한다는 (예를 들어, 대학 캠퍼스의 학생이라는) 것을 검증하기 위해 제2 인증 정도를 요구할 수 있다.

[0080] 위에서 언급된 바와 같이, 비디오 테이블(304)은, 일 예에서, 그에 대해 메시지 테이블(302) 내에 레코드들이 유지되는 메시지들과 연관된 비디오 데이터를 저장한다. 유사하게, 이미지 테이블(312)은 메시지들- 이들에 대한 메시지 데이터가 엔티티 테이블(306)에 저장됨 -과 연관된 이미지 데이터를 저장한다. 엔티티 테이블(306)은 증강 테이블(310)로부터의 다양한 증강들을 이미지 테이블(312) 및 비디오 테이블(304)에 저장되는 다양한 이미지들 및 비디오들과 연관시킬 수 있다.

[0081] 훈련된 머신 러닝 기술(들)(307)은 의도 결정 시스템(224)의 훈련 동안 훈련된 파라미터들을 저장한다. 예를 들어, 훈련된 머신 러닝 기술들(307)은 하나 이상의 신경망 머신 러닝 기술의 훈련된 파라미터들을 저장한다.

[0082] 데이터 통신 아키텍처

[0083] 도 4는 추가 메시징 클라이언트(104) 또는 메시징 서버(118)로의 통신을 위해 메시징 클라이언트(104)에 의해 생성된, 일부 예들에 따른, 메시지(400)의 구조를 예시하는 개략도이다. 특정 메시지(400)의 콘텐츠는 메시징 서버(118)에 의해 액세스가능한, 데이터베이스(126) 내에 저장되는 메시지 테이블(302)을 채우기 위해 사용된다. 유사하게, 메시지(400)의 콘텐츠는 클라이언트 디바이스(102) 또는 애플리케이션 서버들(114)의 "수송 중(in-transit)" 또는 "비행 중(in-flight)" 데이터로서 메모리에 저장된다. 메시지(400)는 이하의 예시적인 컴포넌트들을 포함하는 것으로 도시된다:

[0084] · 메시지 식별자(402): 메시지(400)를 식별하는 고유 식별자

[0085] · 메시지 텍스트 페이로드(404): 클라이언트 디바이스(102)의 사용자 인터페이스를 통해 사용자에 의해 생성되고 메시지(400)에 포함되는 텍스트.

[0086] · 메시지 이미지 페이로드(406): 클라이언트 디바이스(102)의 카메라 컴포넌트에 의해 캡처되거나 클라이언트 디바이스(102)의 메모리 컴포넌트로부터 검색되고, 메시지(400)에 포함되는 이미지 데이터. 송신된 또는 수신된 메시지(400)에 대한 이미지 데이터는 이미지 테이블(312)에 저장될 수 있다.

[0087] · 메시지 비디오 페이로드(408): 카메라 컴포넌트에 의해 캡처되거나 클라이언트 디바이스(102)의 메모리 컴포

넌트로부터 검색되고, 메시지(400)에 포함되는 비디오 데이터. 전송되거나 수신된 메시지(400)에 대한 비디오 데이터가 비디오 테이블(304)에 저장될 수 있다.

[0088] · 메시지 오디오 페이로드(410): 마이크로폰에 의해 캡처되거나 클라이언트 디바이스(102)의 메모리 컴포넌트로부터 검색되고, 메시지(400)에 포함되는 오디오 데이터.

[0089] · 메시지 증강 데이터(412): 메시지(400)의 메시지 이미지 페이로드(406), 메시지 비디오 페이로드(408), 또는 메시지 오디오 페이로드(410)에 적용될 증강들을 나타내는 증강 데이터(예를 들어, 필터들, 스티커들, 또는 다른 주석들 또는 향상들). 전송 또는 수신된 메시지(400)에 대한 증강 데이터는 증강 테이블(310) 내에 저장될 수 있다.

[0090] · 메시지 지속기간 파라미터(414): 메시지의 콘텐츠(예를 들어, 메시지 이미지 페이로드(406), 메시지 비디오 페이로드(408), 메시지 오디오 페이로드(410))가 메시징 클라이언트(104)를 통해 사용자에게 제시되거나 액세스 가능하게 되는 시간의 양을 초 단위로 표시하는 파라미터 값.

[0091] · 메시지 지리위치 파라미터(416): 메시지의 콘텐츠 페이로드와 연관된 지리위치 데이터(예를 들어, 위도 및 경도 좌표들). 다수의 메시지 지리위치 파라미터(416) 값들이 페이로드에 포함될 수 있고, 이들 파라미터 값들 각각은 콘텐츠(예를 들어, 메시지 이미지 페이로드(406) 내의 특정 이미지, 또는 메시지 비디오 페이로드(408) 내의 특정 비디오)에 포함된 콘텐츠 아이템들에 관하여 연관된다.

[0092] · 메시지 스토리 식별자(418): 메시지(400)의 메시지 이미지 페이로드(406) 내의 특정 콘텐츠 아이템이 연관되는 하나 이상의 콘텐츠 컬렉션(예를 들어, 스토리 테이블(314)에서 식별되는 "스토리들")을 식별하는 식별자 값들. 예를 들어, 메시지 이미지 페이로드(406) 내의 다수의 이미지 각각은 식별자 값들을 사용하여 다수의 콘텐츠 컬렉션과 연관될 수 있다.

[0093] · 메시지 태그(420): 각각의 메시지(400)는 다수의 태그들로 태깅될 수 있고, 그 각각은 메시지 페이로드에 포함된 콘텐츠의 주제를 나타낸다. 예를 들어, 메시지 이미지 페이로드(406)에 포함되는 특정 이미지가 동물(예를 들어, 사자)을 묘사하는 경우, 관련 동물을 나타내는 태그 값이 메시지 태그(420) 내에 포함될 수 있다. 태그 값들은 사용자 입력에 기초하여 수동으로 생성될 수 있거나, 예를 들어, 이미지 인식을 사용하여 자동으로 생성될 수 있다.

[0094] · 메시지 발신자 식별자(422): 메시지(400)가 생성되고 그로부터 메시지(400)가 전송된 클라이언트 디바이스(102)의 사용자를 나타내는 식별자(예를 들어, 메시징 시스템 식별자, 이메일 주소, 또는 디바이스 식별자).

[0095] · 메시지 수신자 식별자(424): 메시지(400)가 어드레싱되는 클라이언트 디바이스(102)의 사용자를 나타내는 식별자(예를 들어, 메시징 시스템 식별자, 이메일 주소, 또는 디바이스 식별자).

[0096] 메시지(400)의 다양한 컴포넌트들의 콘텐츠들(예를 들어, 값들)은 콘텐츠 데이터 값들이 저장되는 테이블들 내의 위치들에 대한 포인터들일 수 있다. 예를 들어, 메시지 이미지 페이로드(406)에서의 이미지 값은 이미지 테이블(312) 내의 위치에 대한 포인터(또는 그의 주소)일 수 있다. 유사하게, 메시지 비디오 페이로드(408) 내의 값들은 비디오 테이블(304) 내에 저장되는 데이터에 포인팅할 수 있고, 메시지 증강 데이터(412) 내에 저장되는 값들은 증강 테이블(310)에 저장되는 데이터에 포인팅할 수 있고, 메시지 스토리 식별자(418) 내에 저장되는 값들은 스토리 테이블(314)에 저장되는 데이터에 포인팅할 수 있고, 메시지 발신자 식별자(422) 및 메시지 수신자 식별자(424) 내에 저장되는 값들은 엔티티 테이블(306) 내에 저장되는 사용자 레코드들에 포인팅할 수 있다.

[0097] 의도 결정 시스템

[0098] 도 5는 예시적인 예들에 따른 예시적인 의도 결정 시스템(224)을 도시하는 블록도이다. 의도 결정 시스템(224)은 입력 데이터(예를 들어, 사람 또는 훈련 데이터와 같은 현실 세계 객체를 묘사하는 단안 이미지/비디오)의 세트에 대해 동작하는 컴포넌트들(510)의 세트를 포함한다. 입력 데이터의 세트는 훈련 페이즈들 동안 하나 이상의 데이터베이스(들)(도 3)로부터 획득되고, AR/VR 애플리케이션이, 예컨대 메시징 클라이언트(104)에 의해 사용되고 있을 때 클라이언트 디바이스(102)의 RGB 카메라로부터 획득된다. 의도 결정 시스템(224)은 머신 러닝 기술 모듈(512), 골격 키포인트 모듈(511), 신체 자세 모듈(514), 이미지 수정 모듈(518), AR 효과 모듈(519), 의도 결정 모듈(530), 음성 입력 모듈(532), 3D 신체 추적 모듈(513), 전신 세그멘테이션 모듈(515), 및 이미지 디스플레이 모듈(520)을 포함한다.

[0099] 훈련 동안, 의도 결정 시스템(224)은 훈련 데이터(501)로부터 주어진 훈련 이미지 또는 비디오를 수신한다. 의도 결정 시스템(224)은 주어진 훈련 이미지 또는 비디오에 대해 머신 러닝 기술 모듈(512)을 사용하여 하나 이

상의 머신 러닝 기술을 적용한다. 머신 러닝 기술 모듈(512)은 이미지 또는 비디오에 묘사된 사람(들) 또는 사용자(들)의 3D 신체 메시를 추정하기 위해 주어진 훈련 이미지 또는 비디오로부터 하나 이상의 특징을 추출한다.

[0100] 머신 러닝 기술 모듈(512)은 주어진 훈련 이미지 또는 비디오와 연관된 3D 신체 메시 정보를 검색한다. 머신 러닝 기술 모듈(512)은 추정된 3D 신체 메시를 훈련 데이터(502)의 일부로서 제공된 실측 정보 의복 3D 신체 메시와 비교한다. 비교의 차이 임계값 또는 편차에 기초하여, 머신 러닝 기술 모듈(512)은 하나 이상의 계수 또는 파라미터를 업데이트하고 하나 이상의 추가적인 훈련 이미지 또는 비디오를 획득한다. 훈련 이미지들의 지정된 수의 에포크들 또는 배치들이 처리된 후에 및/또는 차이 임계값 또는 편차가 지정된 값에 도달할 때, 머신 러닝 기술 모듈(512)은 훈련을 완료하고 머신 러닝 기술 모듈(512)의 파라미터들 및 계수들은 훈련된 머신 러닝 기술(들)(307)에 저장된다.

[0101] 일부 예들에서, 훈련 동안, 머신 러닝 기술 모듈(512)은 골격 키포인트 모듈(511)로부터 2D 골격 관절 정보를 수신한다. 골격 키포인트 모듈(511)은 주어진 훈련 이미지에 묘사된 사용자의 골격 키포인트들(예를 들어, 머리 관절, 어깨 관절들, 고관절들, 다리 관절들 등)을 추적하고 골격 키포인트들의 2D 또는 3D 좌표들을 제공한다. 이 정보는 머신 러닝 기술 모듈(512)에 의해 훈련 이미지에 묘사된 사용자의 구별 속성들을 식별하고 3D 신체 메시를 생성하는 데 사용된다.

[0102] 머신 러닝 기술 모듈(512)에 의해 생성된 3D 신체 메시는 신체 자세 모듈(514)에 제공된다. 신체 자세 모듈(514)은 이미지 또는 비디오에 묘사된 객체를 추적하고 객체와 연관된 3D 신체 메시를 업데이트할 수 있다. 예에서, 신체 자세 모듈(514)은 3D 신체 추적 모듈(513)에 의해 제공되는 3D 신체 추적 정보에 기초하여 객체를 추적할 수 있다. 신체 자세 모듈(514)은 일부 예들에 따라, 이미지 또는 비디오에 묘사된 객체의 자세를 결정할 수 있다. 도 6은 예시적인 실시예에 따른 의도 결정 시스템(224)의 출력들의 도식적 표현이다. 구체적으로, 도 6은 신체 자세 모듈(514)에 의해 생성되고 추적되는 3D 신체 메시(600)를 도시한다. 일 예에서, 신체 자세 모듈(514)은 비디오의 프레임들에 걸쳐 3D 신체 메시의 변화들을 추적할 수 있다. 신체 자세 모듈(514)은 3D 신체 메시가 외부 메시에 혼합됨에 따라 3D 신체 메시에 대한 변경들에 기초하여 외부 메시를 업데이트하고 변형시키기 위해 3D 신체 메시에 대한 변경들을 의도 결정 모듈(530)에 제공할 수 있다.

[0103] 예에서, 신체 자세 모듈(514)은 이미지 또는 비디오에 묘사된 객체에 대응하는 주어진 자세 또는 신체 메시를 타겟 자세와 매칭시킬 수 있다. 신체 자세 모듈(514)은 이미지 또는 비디오에 묘사된 객체의 신체 메시 또는 골격 키포인트들을 복수의 상이한 타겟 자세와 비교할 수 있다. 복수의 상이한 타겟 자세들 각각은 음성 입력을 명확하게 하는 데 사용될 수 있는 상이한 의도와 연관될 수 있다. 예를 들어, 제1 타겟 자세는 특정 방향을 가리키는 손가락 또는 손으로 인체를 묘사할 수 있다. 이러한 경우들에서, 제1 타겟 자세와 연관된 의도는 특정 방향의 지정된 범위 내에 있는 객체, 대상, 아이템 또는 방향인 것으로 식별될 수 있다. 아래에 설명되는 바와 같이, 가상선은 손가락 또는 손으로부터 특정 방향을 따라 연장될 수 있다. 가상선과 교차하는 제1 객체는 예컨대 대명사 또는 부사를 제1 객체의 신원으로 대체함으로써 의도를 분별하거나 명확하게 하기 위해 사용될 수 있다. 다른 예로서, 제2 타겟 자세는 인간이 그들의 입술 또는 다리를 사용하는 것과 같은 제스처를 수행하는 것을 묘사할 수 있다. 제스처는 특정 방향으로 배향되거나 포커싱될 수 있다. 이러한 경우들에서, 제2 타겟 자세와 연관된 의도는 특정 방향의 지정된 범위 내에 있는 객체, 대상, 아이템 또는 방향인 것으로 식별될 수 있다. 다시, 가상선은 특정 방향을 따라 입술들 또는 다리들로부터 연장될 수 있다. 가상선과 교차하는 제2 객체는 예컨대 대명사 또는 부사를 제2 객체의 신원으로 대체함으로써 의도를 분별하거나 명확하게 하기 위해 사용될 수 있다.

[0104] 신체 자세 모듈(514)에 의해 결정된 자세와 연관된 의도는 의도 결정 모듈(530)에 제공될 수 있다. 예를 들어, 신체 자세 모듈(514)은 자세가 사람이 특정 방향을 향해 가리키거나 제스처를 취하는 것을 표현한다는 것을 의도 결정 모듈(530)에 표시할 수 있다. 의도 결정 모듈(530)은 음성 입력 모듈(532)과 통신하여 이미지 또는 비디오에 묘사된 사람으로부터 구두 커맨드를 수신할 수 있다. 음성 입력 모듈(532)은 구두 커맨드에 포함되는 대명사 또는 부사와 같은 구두 커맨드의 모호한 부분을 식별할 수 있다. 일부 경우들에서, 신체 자세 모듈(514)은 이미지 또는 비디오에 묘사된 사람에 대해 결정된 자세에 기초하여 생성되는 가상선을 제공할 수 있다. 즉, 신체 자세 모듈(514)은 주어진 방향을 따라 신체 메시의 일부로부터 연장되는 가상선을 생성할 수 있다. 의도 결정 모듈(530)은 가상선을 사용하여 이미지 또는 비디오에서 가상선과 교차하는 하나 이상의 객체를 식별하여 구두 커맨드의 모호한 의도를 해석한다.

[0105] 의도 결정 모듈(530)은 신체 자세 모듈(514)로부터 수신된 자세와 연관된 의도에 기초하여 구두 커맨드의 의도

를 명확하게 할 수 있다. 예를 들어, 의도 결정 모듈(530)은 구두 커맨드의 모호한 부분이 이미지 또는 비디오에 묘사된 객체, 아이템, 또는 요소를 지칭한다고 결정할 수 있다. 구체적으로, 구두 커맨드는 "이것을 청색으로 변경"할 수 있다. 이러한 상황들에서, 단어 "이것"은 알려지지 않은 객체, 아이템 또는 요소를 지칭하는 대명사인 것으로 결정될 수 있다. 의도 결정 모듈(530)은 신체 자세 모듈(514)로부터 수신된 자세와 연관된 의도를 처리하여 단어 "이것"에 의해 지칭되는 객체, 아이템 또는 요소를 식별할 수 있다. 즉, 의도 결정 모듈(530)은 이미지 또는 비디오에 묘사된 사람에 대응하는 신체 자세의 손가락, 손, 또는 제스처로부터 연장되는 가상선과 교차하는 객체, 아이템, 대상, 또는 요소를 식별할 수 있다. 의도 결정 모듈(530)은 대명사 "이것"을 식별된 객체, 아이템, 대상 또는 요소로 대체할 수 있다. 예를 들어, 커맨드의 화자가 착용한 모자는 손가락 또는 손으로부터 손가락 또는 손이 가리키는 방향을 향해 그려진 가상선과 교차할 수 있다. 그 결과, 의도 결정 모듈(530)은 음성 커맨드가 모자를 청색으로 변경하도록(예를 들어, 모자의 속성을 변경하도록) 요청하고 있다고 결정할 수 있다.

[0106] 예를 들어, 도 7에 도시된 바와 같이, 이미지 또는 비디오(700)는 사용자(701)가 주어진 자세를 수행하는 것을 묘사한다. 신체 자세 모듈(514)은 사용자(701)의 일부로부터 특정 방향을 향해 연장되는 방향 또는 가상선(710)을 제공할 수 있다. 의도 결정 모듈(530)은 "이것을 청색으로 만들어줘"와 같은 구두 커맨드를 수신하고, 단어 "이것"이 대명사가 지칭하는 아이템, 객체, 대상 또는 사물에 관해 모호하다고 결정할 수 있다. 의도 결정 모듈(530)은 가상선(710)과 교차하는 이미지 또는 비디오(700)에 묘사되는 현실 세계 아이템을 식별하기 위해 이미지 또는 비디오(700)에 가상선(710)을 적용할 수 있다. 이 경우, 의도 결정 모듈(530)은 사용자(701)가 착용한 모자(720)를 가상선(710)과 교차하는 것으로서 식별한다. 그 결과, 의도 결정 모듈(530)은 구두 커맨드가 모자(720)의 속성(예를 들어, 모자(720)의 색상)을 청색으로 변경하도록 요청한다고 결정한다. 이에 응답하여, 의도 결정 모듈(530)은 AR 효과 모듈(519)과 통신하여 하나 이상의 AR 요소를 모자(720)에 적용하여 현실 세계 모자(720)의 픽셀 색상들을 청색으로 변경한다. 예로서, AR 효과 모듈(519)은 현실 세계 모자를 청색의 가상 모자(720)로 대체하기 위해 현실 세계 모자(720)를 포함하는 이미지 또는 비디오(700)의 부분에 가상 모자를 적용한다. 스타일, 크기, 패턴, 타입 등과 같은 다른 속성들이 음성 입력에 기초하여 유사하게 변경될 수 있다. 예를 들어, "이것을 청색으로 만들어줘"(색상 속성을 지칭함)라고 말하는 대신에, 음성 입력은 "이것을 더 길게 만들어줘"(크기 속성의 변화를 지칭함), "이것을 물방울 무늬로 만들어줘"(패턴 속성의 변화를 지칭함), 및/또는 "이것을 여름에 어울리게 만들어줘"(스타일의 변화를 지칭함)라고 말할 수 있다.

[0107] 예에서, AR 효과 모듈(519)은 이미지 또는 비디오에 묘사된 객체와 연관된 신체 메시에 기초하여 그리고 의도 결정 모듈(530)에 의해 제공된 해석된 음성 커맨드에 기초하여 이미지 또는 비디오에 묘사된 객체에 하나 이상의 AR 요소 또는 그래픽을 선택하고 적용한다. 이미지 또는 비디오에 묘사된 현실 세계 객체와 조합된 이 AR 그래픽은 AR 지갑 또는 귀걸이와 같은 AR 객체를 착용한 사람을 묘사하는 이미지 또는 비디오를 렌더링하기 위해 이미지 수정 모듈(518)에 제공된다. 일 예에서, AR 효과 모듈(519)은 의도 결정 모듈(530)로부터 수신된 커맨드에 기초하여 복수의 패션 아이템 세그멘테이션 모듈들 사이에서 선택할 수 있다. 즉, 의도 결정 모듈(530)이 모호한 의도를 포함하는 구두 커맨드가 모자를 지칭하고 있다는 것을 AR 효과 모듈(519)에 표시하는 경우, AR 효과 모듈(519)은 모자 세그멘테이션 모듈을 선택한다. 모자 세그멘테이션 모듈은 모자 타입의 패션 아이템의 경계들을 식별하고 정의하기 위해 이미지를 처리하도록 훈련된다. AR 효과 모듈(519)은 이어서 모자 세그멘테이션 모듈의 세그멘테이션 출력 내의 픽셀들을 가상 객체의 픽셀들로 대체할 수 있다. AR 효과 모듈(519)은 또한 모자 세그멘테이션 모듈의 세그멘테이션 출력에 기초하여 가상 객체의 3D 위치를 추적하고 수정한다. 다른 예로서, 의도 결정 모듈(530)이, 모호한 의도를 포함하는 구두 커맨드가 상의 또는 셔츠 패션 아이템을 지칭하고 있다는 것을 AR 효과 모듈(519)에 표시하는 경우, AR 효과 모듈(519)은 상의 세그멘테이션 모듈을 선택한다. 상의 세그멘테이션 모듈은 블라우스, 셔츠, 탱크 탑, 스웨터 등과 같은 상의 타입의 패션 아이템의 경계들을 식별하고 정의하기 위해 이미지를 처리하도록 훈련된다. AR 효과 모듈(519)은 이어서 상의 세그멘테이션 모듈의 세그멘테이션 출력 내의 픽셀들을 가상 객체의 픽셀들로 대체할 수 있다. AR 효과 모듈(519)은 또한 상의 세그멘테이션 모듈의 세그멘테이션 출력에 기초하여 가상 객체의 3D 위치를 추적하고 수정한다.

[0108] 이미지 수정 모듈(518)은 AR 효과 모듈(519)에 의해 선택된 AR 효과에 기초하여 카메라에 의해 캡처된 이미지를 조정할 수 있다. 이미지 수정 모듈(518)은, 예컨대 사용자의 3D 신체 메시에 대한 변경들에 기초하여 AR 의복 또는 패션 액세서리의 물리적 특성들(변형)을 변경함으로써, 이미지 또는 비디오에 묘사된 사용자 또는 사람 위에 배치된 AR 의복(들) 또는 패션 액세서리가 이미지 또는 비디오에 제시되는 방식을 조정한다. 이미지 디스플레이 모듈(520)은 이미지 수정 모듈(518)에 의해 이루어진 조정들을 사용자의 신체를 묘사하는 수신된 단안 이미지 또는 비디오로 조합한다. 이미지 또는 비디오는 이미지 디스플레이 모듈(520)에 의해 클라이언트 디바이

스(102)에 제공되고, 이어서 다른 사용자에게 전송되거나 나중의 액세스 및 디스플레이를 위해 저장될 수 있다.

[0109] 일부 예들에서, 이미지 수정 모듈(518)은 3D 신체 추적 모듈(513)로부터 이미지에 묘사된 사용자의 3D 위치들을 나타내는 3D 신체 추적 정보를 수신할 수 있다. 3D 신체 추적 모듈(513)은 추가적인 머신 러닝 기술들을 사용하여 훈련 데이터(501)를 처리함으로써 3D 신체 추적 정보를 생성한다. 이미지 수정 모듈(518)은 또한 이미지 내의 어느 픽셀들이 사용자의 전신에 대응하는지를 나타내는 전신 세그멘테이션을 다른 머신 러닝 기술로부터 수신할 수 있다. 전신 세그멘테이션은 전신 세그멘테이션 모듈(515)로부터 수신될 수 있다. 전신 세그멘테이션 모듈(515)은 머신 러닝 기술을 사용하여 훈련 데이터(501)를 처리함으로써 전신 세그멘테이션을 생성한다.

[0110] 신체 자세 모듈(514)에 의해 결정된 자세와 연관된 의도는 의도 결정 모듈(530)에 제공될 수 있다. 예를 들어, 신체 자세 모듈(514)은 자세가 사람이 특정 방향을 향해 가리키거나 제스처를 취하는 것을 표현한다는 것을 의도 결정 모듈(530)에 표시할 수 있다. 의도 결정 모듈(530)은 음성 입력 모듈(532)과 통신하여 이미지 또는 비디오에 묘사된 사람으로부터 구두 커맨드를 수신할 수 있다. 음성 입력 모듈(532)은 구두 커맨드에 포함되는 대명사 또는 부사와 같은 구두 커맨드의 모호한 부분을 식별할 수 있다. 일부 경우들에서, 신체 자세 모듈(514)은 이미지 또는 비디오에 묘사된 사람에 대해 결정된 자세에 기초하여 생성되는 가상선을 제공할 수 있다. 즉, 신체 자세 모듈(514)은 주어진 방향을 따라 신체 메시의 일부로부터 연장되는 가상선을 생성할 수 있다. 의도 결정 모듈(530)은 가상선을 사용하여 이미지 또는 비디오에서 가상선과 교차하는 하나 이상의 객체를 식별하여 구두 커맨드의 모호한 의도를 해석한다.

[0111] 다른 예로서, 의도 결정 모듈(530)은 구두 커맨드의 모호한 부분이 커맨드의 화자를 묘사하는 이미지 또는 비디오의 부분의 모호한 색상(또는 다른 속성)를 지칭한다고 결정할 수 있다. 구체적으로, 구두 커맨드는 "모자를 이 색상으로 바꿔줘"일 수 있다. 이러한 상황들에서, 단어 "이"는 미지의 색상 값 또는 속성을 지칭하는 대명사인 것으로 결정될 수 있다. 의도 결정 모듈(530)은 신체 자세 모듈(514)로부터 수신된 자세와 연관된 의도를 처리하여 단어 "이"가 지칭하는 색상 또는 속성을 식별할 수 있다. 즉, 의도 결정 모듈(530)은 이미지 또는 비디오에 묘사된 사람에 대응하는 신체 자세의 손가락, 손, 또는 제스처로부터 연장되는 가상선과 교차하는 객체, 아이템, 대상, 또는 요소를 식별할 수 있다. 의도 결정 모듈(530)은 대명사 "이"를 식별된 객체, 아이템, 대상 또는 요소의 색상 또는 속성으로 대체할 수 있다. 예를 들어, 손가락 또는 손으로부터 손가락 또는 손이 가리키는 방향을 향해 그려진 가상선이 적색인 다른 묘사된 객체(예를 들어, 현실 세계 벽, 하늘, 물, 책상, 램프 등)와 교차하는 경우, 커맨드의 화자가 착용한 모자가 적색으로 변경될 수 있다. 그 결과, 의도 결정 모듈(530)은 음성 커맨드가 모자를 적색으로 변경하도록 요청하고 있다고 결정할 수 있다. 일부 경우들에서, 의도 결정 모듈(530)은 가상선과 교차하는 어떠한 객체도 식별하지 못한다. 이러한 경우들에서, 의도 결정 모듈(530)은 화자의 손가락의 임계 거리 내에 있는 픽셀의 색상을 검색할 수 있다.

[0112] 다른 예로서, 의도 결정 모듈(530)은 구두 커맨드의 모호한 부분이 모호한 방향을 지칭한다고 결정할 수 있다. 구체적으로, 구두 커맨드는 "저기로 공을 던져"일 수 있다. 이러한 상황들에서, 단어 "저기"는 알려지지 않은 방향을 지칭하는 부사인 것으로 결정될 수 있다. 의도 결정 모듈(530)은 신체 자세 모듈(514)로부터 수신된 자세와 연관된 의도를 처리하여 단어 "저기"가 지칭하는 방향을 식별할 수 있다. 즉, 의도 결정 모듈(530)은 이미지 또는 비디오에 묘사된 사람에 대응하는 신체 자세의 손가락, 손, 또는 제스처로부터 연장되는 가상선을 획득할 수 있다. 의도 결정 모듈(530)은 부사 "저기"를 가상선의 방향으로 대체할 수 있다. 그 결과, 의도 결정 모듈(530)은 가상 객체 또는 공을 가상선의 방향을 따라 던지거나 이동하는 것으로 애니메이션하도록 AR 효과 모듈(519)에 지시할 수 있다.

[0113] 다른 예로서, 의도 결정 모듈(530)은 구두 커맨드의 모호한 부분이 커맨드의 화자를 묘사하는 이미지 또는 비디오의 부분에 묘사된 모호한 사람을 지칭한다고 결정할 수 있다. 즉, 이미지 또는 비디오는 다수의 사람들을 묘사할 수 있고, 구두 커맨드는 "내가 저 사람의 셔츠를 입은 모습을 보여줘"일 수 있다. 이러한 상황들에서, 단어 "저 사람"은 미지의 인물을 지칭하는 대명사인 것으로 결정될 수 있다. 의도 결정 모듈(530)은 단어 "저 사람"이 지칭하는 사람을 식별하기 위해 신체 자세 모듈(514)로부터 수신된 자세와 연관된 의도를 처리할 수 있다. 즉, 의도 결정 모듈(530)은 이미지 또는 비디오에 묘사된 사람(예를 들어, 커맨드를 말하는 제1 인물)에 대응하는 신체 자세의 손가락, 손, 또는 제스처로부터 연장되는 가상선과 교차하는 다른 사람(이미지 또는 비디오에 묘사된 제2 인물)을 식별할 수 있다. 의도 결정 모듈(530)은 대명사 "저 사람"을 제2 인물으로 대체할 수 있다.

[0114] 이에 응답하여, 의도 결정 모듈(530)은 명확한 인물(예를 들어, 제2 인물)이 착용한 구두 커맨드에서 식별된 패션 아이템(예를 들어, 셔츠)을 검색하도록 AR 효과 모듈(519)에 지시할 수 있다. 의도 결정 모듈(530)은 제2

인물의 UV 또는 3D 좌표들을 AR 효과 모듈(519)에 제공할 수 있다. 이에 응답하여, AR 효과 모듈(519)은 셔츠에 대응하는 세그멘테이션 모듈, 이 경우에는 상의 세그멘테이션 모듈을 획득할 수 있다. AR 효과 모듈(519)은 이미지에 묘사된 제2 인물에게 상의 세그멘테이션 모듈을 적용하여 제2 인물이 착용한 셔츠의 세그멘테이션을 획득할 수 있다. AR 효과 모듈(519)은 이어서, 셔츠 세그멘테이션에 대응하고 제2 인물이 착용한 셔츠의 픽셀 값들을 포함하는 AR 형상을 생성함으로써 제2 인물이 착용한 셔츠를 표현하는 AR 객체를 생성할 수 있다. AR 효과 모듈(519)은 이어서, 예컨대, 제1 인물의 상체에 맞도록 AR 객체를 재스케일링하고 크기 조정함으로써, 생성된 AR 객체를 제1 인물에게 적용할 수 있다. "내가 저 사람의 셔츠를 착용한 모습을 보여줘"라는 음성 커맨드를 수신한 결과로서, AR 효과 모듈(519)은 이미지 또는 비디오에 묘사된 다른 인물이 착용한 셔츠를 커맨드의 화자가 착용한 것으로서 렌더링할 수 있다.

[0115] 도 8에 도시된 바와 같이, 이미지(800)는 클라이언트 디바이스(102)의 카메라에 의해 캡처된 제1 현실 세계 인물(810) 및 제2 현실 세계 인물(820)을 묘사한다. 음성 또는 음성 입력은 제1 현실 세계 인물(810)(클라이언트 디바이스(102)의 사용자)으로부터 수신될 수 있다. 음성 또는 음성 입력은 "내가 저 사람의 셔츠를 입은 모습을 보여줘"일 수 있다. 의도 결정 모듈(530)은 단어 "저 사람"이 음성 또는 음성 입력이 모호한 의도를 갖게 하는 대명사라고 결정할 수 있다. 의도 결정 모듈(530)은 제1 현실 세계 인물(810)의 자세(제1 현실 세계 인물(810)의 손가락이 제2 현실 세계 인물(820)의 방향을 가리킴)에 기초하여 단어 "저 사람"이 제2 현실 세계 인물(820)(음성 또는 음성 입력의 비-화자)을 지칭한다는 것을 분별할 수 있다. 이에 응답하여, 의도 결정 모듈(530)은 제2 현실 세계 인물(820)의 UV 또는 3D 좌표들을 AR 효과 모듈(519) 및 제1 현실 세계 인물(810) 상에 렌더링될 필요가 있는 타겟 현실 세계 객체(830)(예를 들어, 셔츠)의 식별을 제공할 수 있다. 이에 응답하여, AR 효과 모듈(519)은 타겟 현실 세계 객체(830)의 세그멘테이션을 획득할 수 있고, 그 후 타겟 현실 세계 객체(830)의 생성된 AR 객체를 제1 현실 세계 인물(810)에 적용할 수 있다.

[0116] 일 구현에서, 제2 현실 세계 인물(820)이 착용한 것으로 묘사된 타겟 현실 세계 객체(830)를 제1 현실 세계 인물(810)에 착용되는 것으로 렌더링하기 위해, AR 효과 모듈(519)은 제1 및 제2 현실 세계 인물들(810 및 820)의 대응하는 부분들의 제1 및 제2 세그멘테이션들을 생성한다. 즉, AR 효과 모듈(519)은(제2 현실 세계 인물(820)을 지칭하는 것으로 명확화된) 제1 현실 세계 인물(810)로부터 음성 입력을 수신하는 것에 응답하여 패션 아이템의 제1 세그멘테이션을 생성하기 위해 제2 현실 세계 인물(820)이 착용한 패션 아이템(예를 들어, 셔츠 또는 타겟 현실 세계 객체(830))을 세그멘테이션한다. AR 효과 모듈(519)은 음성 입력을 수신하는 것에 응답하여 패션 아이템의 제2 세그멘테이션을 생성하기 위해 제1 현실 세계 인물(810)이 착용한 (예를 들어, 타겟 현실 세계 객체(830)에 대응하는) 패션 아이템(예를 들어, 제1 현실 세계 인물(810)이 착용한 셔츠)을 세그멘테이션한다. AR 효과 모듈(519)은 제1 세그멘테이션에 기초하여 제2 현실 세계 인물(820)이 착용한 패션 아이템의 픽셀 값들을 획득하고, 패션 아이템의 제2 세그멘테이션 및 획득된 픽셀 값들에 기초하여 제1 현실 세계 인물(810)이 착용한 패션 아이템에 하나 이상의 AR 요소를 적용한다. 즉, AR 효과 모듈(519)은 제2 현실 세계 인물(820)이 착용한 패션 아이템의 제1 세그멘테이션 내의 픽셀 값들에 기초하여(매칭하거나 미러링하기 위해) 제1 현실 세계 인물(810)이 착용한 패션 아이템의 제2 세그멘테이션 내의 픽셀 값들을 변색시킬 수 있다.

[0117] 도 9는 일부 예들에 따른, 의도 결정 시스템(224)에 의해 수행되는 프로세스(900)의 흐름도이다. 흐름도는 동작들을 순차적 프로세스로서 설명할 수 있지만, 동작들 중 다수는 병렬로 또는 동시에 수행될 수 있다. 또한, 동작들의 순서는 재배열될 수 있다. 프로세스는 그 동작들이 완료될 때 종료된다. 프로세스는 방법, 절차 등에 대응할 수 있다. 방법들의 단계들은 전체적으로 또는 부분적으로 수행될 수 있고, 다른 방법들에서의 단계들의 일부 또는 전부와 함께 수행될 수 있고, 임의의 수의 상이한 시스템들 또는 시스템들 중 임의의 것에 포함된 프로세서와 같은 그 임의의 부분에 의해 수행될 수 있다.

[0118] 동작 901에서, 의도 결정 시스템(224)(예를 들어, 클라이언트 디바이스(102) 또는 서버)은 위에서 논의된 바와 같이 사람을 묘사하는 이미지 또는 비디오를 수신한다.

[0119] 동작 902에서, 의도 결정 시스템(224)은 위에서 논의된 바와 같이, 사람의 골격 관절들의 세트를 식별한다.

[0120] 동작 903에서, 의도 결정 시스템(224)은 위에서 논의된 바와 같이, 골격 관절들의 세트의 위치확인에 기초하여 이미지에 묘사된 사람의 자세를 식별한다.

[0121] 동작 904에서, 의도 결정 시스템(224)은 증강 현실(AR) 동작을 수행하라는 요청을 포함하는 음성 입력을 수신하고, 음성 입력은 위에서 논의된 바와 같이 모호한 의도를 포함한다.

[0122] 동작 905에서, 의도 결정 시스템(224)은 위에서 논의된 바와 같이, 이미지에 묘사된 사람의 자세에 기초하여 음

성 입력의 모호한 의도를 식별한다.

[0123] 동작 906에서, 의도 결정 시스템(224)은, 음성 입력을 수신하는 것에 응답하여, 위에서 논의된 바와 같이, 이미지에 묘사된 사람의 자세에 기초하여 음성 입력의 모호한 의도를 분별하는 것에 기초하여 AR 동작을 수행한다.

[0124] 머신 아키텍처

[0125] 도 10은 머신(1000)으로 하여금 본 명세서에서 논의된 방법론들 중 어느 하나 이상을 수행하게 하기 위한 명령어들(1008)(예를 들어, 소프트웨어, 프로그램, 애플리케이션, 애플릿, 앱, 또는 다른 실행가능 코드)이 실행될 수 있는 머신(1000)의 도식적 표현이다. 예를 들어, 명령어들(1008)은 머신(1000)으로 하여금 본 명세서에 설명된 방법들 중 어느 하나 이상을 실행하게 할 수 있다. 명령어들(1008)은 범용의, 프로그램되지 않은 머신(1000)을 설명되고 예시된 기능들을 설명된 방식으로 수행하도록 프로그램된 특정 머신(1000)으로 변환한다. 머신(1000)은 독립형 디바이스로서 동작할 수 있거나 다른 머신들에 결합(예를 들어, 네트워크화)될 수 있다. 네트워크화된 배치에서, 머신(1000)은 서버-클라이언트 네트워크 환경에서 서버 머신 또는 클라이언트 머신의 능력으로, 또는 피어-투-피어(또는 분산형) 네트워크 환경에서 피어 머신으로서 동작할 수 있다. 머신(1000)은 서버 컴퓨터, 클라이언트 컴퓨터, 개인용 컴퓨터(PC), 태블릿 컴퓨터, 랩톱 컴퓨터, 넷북, 셋톱 박스(STB), 개인 휴대 정보 단말기(PDA), 엔터테인먼트 미디어 시스템, 셀룰러 전화, 스마트폰, 모바일 디바이스, 웨어러블 디바이스(예를 들어, 스마트워치), 스마트 홈 디바이스(예를 들어, 스마트 어플라이언스), 다른 스마트 디바이스들, 웹 어플라이언스, 네트워크 라우터, 네트워크 스위치, 네트워크 브리지, 또는 머신(1000)에 의해 취해질 액션들을 특정하는 명령어들(1008)을 순차적으로 또는 다른 방식으로 실행할 수 있는 임의의 머신을 포함할 수 있지만, 이에 제한되지 않는다. 또한, 단일 머신(1000)만이 예시되어 있지만, "머신"이라는 용어는 또한 본 명세서에서 논의된 방법론들 중 어느 하나 이상을 수행하기 위해 명령어들(1008)을 개별적으로 또는 공동으로 실행하는 머신들의 컬렉션을 포함하는 것으로 고려되어야 한다. 예를 들어, 머신(1000)은 클라이언트 디바이스(102) 또는 메시징 서버 시스템(108)의 일부를 형성하는 다수의 서버 디바이스들 중 어느 하나를 포함할 수 있다. 일부 예들에서, 머신(1000)은 또한, 특정 방법 또는 알고리즘의 특정 동작들은 서버측에서 수행되고 특정 방법 또는 알고리즘의 특정 동작들은 클라이언트측에서 수행되는 상태로, 클라이언트 및 서버 시스템들 둘 다를 포함할 수 있다.

[0126] 머신(1000)은 버스(1040)를 통해 서로 통신하도록 구성될 수 있는 프로세서들(1002), 메모리(1004), 및 입력/출력(I/O) 컴포넌트들(1038)을 포함할 수 있다. 예에서, 프로세서들(1002)(예를 들어, CPU(Central Processing Unit), RISC(Reduced Instruction Set Computing) 프로세서, CISC(Complex Instruction Set Computing) 프로세서, GPU(Graphics Processing Unit), DSP(Digital Signal Processor), ASIC(Application Specific Integrated Circuit), RFIC(Radio-Frequency Integrated Circuit), 다른 프로세서, 또는 이들의 임의의 적절한 조합)은, 예를 들어, 명령어들(1008)을 실행하는 프로세서(1006) 및 프로세서(1010)를 포함할 수 있다. "프로세서"라는 용어는 명령어들을 동시에 실행할 수 있는 2개 이상의 독립적인 프로세서들(때때로 "코어들"이라고 지칭됨)을 포함할 수 있는 멀티-코어 프로세서들을 포함하도록 의도된다. 도 10은 다수의 프로세서들(1002)을 도시하지만, 머신(1000)은 단일 코어를 갖는 단일 프로세서, 다수의 코어들을 갖는 단일 프로세서(예를 들어, 멀티 코어 프로세서), 단일 코어를 갖는 다수의 프로세서들, 다수의 코어들을 갖는 다수의 프로세서들, 또는 이들의 임의의 조합을 포함할 수 있다.

[0127] 메모리(1004)는 메인 메모리(1012), 정적 메모리(1014), 및 저장 유닛(1016)을 포함하며, 이들 모두는 버스(1040)를 통해 프로세서들(1002)에 액세스가능하다. 메인 메모리(1004), 정적 메모리(1014), 및 저장 유닛(1016)은 본 명세서에 설명된 방법론들 또는 기능들 중 어느 하나 이상을 구현하는 명령어들(1008)을 저장한다. 명령어들(1008)은 또한, 머신(1000)에 의한 그의 실행 동안, 완전히 또는 부분적으로, 메인 메모리(1012) 내에, 정적 메모리(1014) 내에, 저장 유닛(1016) 내의 머신 판독가능 매체(1018) 내에, 프로세서들(1002) 중 적어도 하나 내에(예를 들어, 프로세서의 캐시 메모리 내에), 또는 이들의 임의의 적절한 조합으로 상주할 수 있다.

[0128] I/O 컴포넌트들(1038)은 입력을 수신하고, 출력을 제공하고, 출력을 생성하고, 정보를 송신하고, 정보를 교환하고, 측정들을 캡처하는 등을 위한 매우 다양한 컴포넌트들을 포함할 수 있다. 특정 머신에 포함되는 특정 I/O 컴포넌트들(1038)은 머신의 타입에 의존할 것이다. 예를 들어, 이동 전화들과 같은 휴대용 머신들은 터치 입력 디바이스 또는 다른 이러한 입력 메커니즘들을 포함할 수 있는 반면, 헤드리스 서버 머신은 이러한 터치 입력 디바이스를 포함하지 않을 가능성이 있을 것이다. I/O 컴포넌트들(1038)은 도 10에 도시되지 않은 많은 다른 컴포넌트들을 포함할 수 있다는 것을 이해할 것이다. 다양한 예들에서, I/O 컴포넌트들(1038)은 사용자 출력 컴포넌트들(1024) 및 사용자 입력 컴포넌트들(1026)을 포함할 수 있다. 사용자 출력 컴포넌트들(1024)은 시각

적 컴포넌트들(예를 들어, 플라즈마 디스플레이 패널(PDP), 발광 다이오드(LED) 디스플레이, 액정 디스플레이(LCD), 프로젝터, 또는 음극선관(CRT)과 같은 디스플레이), 음향 컴포넌트들(예를 들어, 스피커들), 햅틱 컴포넌트들(예를 들어, 진동 모터, 저항 메커니즘들), 다른 신호 생성기들 등을 포함할 수 있다. 사용자 입력 컴포넌트들(1026)은 영숫자 입력 컴포넌트들(예를 들어, 키보드, 영숫자 입력을 수신하도록 구성된 터치 스크린, 포토-광학 키보드, 또는 다른 영숫자 입력 컴포넌트들), 포인트 기반 입력 컴포넌트들(예를 들어, 마우스, 터치패드, 트랙볼, 조이스틱, 모션 센서, 또는 다른 포인팅 기구), 촉각 입력 컴포넌트들(예를 들어, 물리적 버튼, 터치들 또는 터치 제스처들의 위치 및 힘을 제공하는 터치 스크린, 또는 다른 촉각 입력 컴포넌트들), 오디오 입력 컴포넌트들(예를 들어, 마이크로폰) 등을 포함할 수 있다.

[0129] 추가 예들에서, I/O 컴포넌트들(1038)은, 광범위한 다른 컴포넌트들 중에서도, 바이오메트릭 컴포넌트들(1028), 모션 컴포넌트들(1030), 환경 컴포넌트들(1032), 또는 위치 컴포넌트들(1034)을 포함할 수 있다. 예를 들어, 바이오메트릭 컴포넌트들(1028)은 표현들(예를 들어, 손 표현들, 얼굴 표정들, 음성 표현들, 신체 제스처들, 또는 눈 추적)을 검출하고, 생체신호들(예를 들어, 혈압, 심박수, 체온, 땀, 또는 뇌파들)을 측정하고, 사람을 식별(예를 들어, 음성 식별, 망막 식별, 얼굴 식별, 지문 식별, 또는 뇌전도 기반 식별) 등을 수행하는 컴포넌트들을 포함한다. 모션 컴포넌트들(1030)은 가속도 센서 컴포넌트들(예를 들어, 가속도계), 중력 센서 컴포넌트들, 회전 센서 컴포넌트들(예를 들어, 자이로스코프)을 포함한다.

[0130] 환경 컴포넌트들(1032)은, 예를 들어, (정지 이미지/사진 및 비디오 능력들을 갖는) 하나 이상의 카메라, 조명 센서 컴포넌트들(예를 들어, 광도계), 온도 센서 컴포넌트들(예를 들어, 주변 온도를 검출하는 하나 이상의 온도계), 습도 센서 컴포넌트들, 압력 센서 컴포넌트들(예를 들어, 기압계), 음향 센서 컴포넌트들(예를 들어, 배경 잡음을 검출하는 하나 이상의 마이크로폰), 근접 센서 컴포넌트들(예를 들어, 인근 객체들을 검출하는 적외선 센서들), 가스 센서들(예를 들어, 안전을 위해 유해 가스들의 농도들을 검출하거나 대기 내의 오염물질들을 측정하기 위한 가스 검출 센서들), 또는 주변 물리적 환경에 대응하는 표시들, 측정들, 또는 신호들을 제공할 수 있는 다른 컴포넌트들을 포함한다.

[0131] 카메라들과 관련하여, 클라이언트 디바이스(102)는 예를 들어, 클라이언트 디바이스(102)의 전방 표면 상의 전면 카메라들 및 클라이언트 디바이스(102)의 후방 표면 상의 후면 카메라들을 포함하는 카메라 시스템을 가질 수 있다. 전면 카메라들은, 예를 들어, 클라이언트 디바이스(102)의 사용자의 정지 이미지들 및 비디오(예를 들어, "셀카들")를 캡처하는 데 사용될 수 있으며, 이는 이어서 위에서 설명한 증강 데이터(예를 들어, 필터들)로 증강될 수 있다. 후면 카메라들은, 예를 들어, 보다 전통적인 카메라 모드에서 정지 이미지들 및 비디오들을 캡처하는 데 사용될 수 있으며, 이러한 이미지들은 유사하게 증강 데이터로 증강된다. 전방 및 후면 카메라들 외에도, 클라이언트 디바이스(102)는 360° 사진들 및 비디오들을 캡처하기 위한 360° 카메라를 또한 포함할 수 있다.

[0132] 추가로, 클라이언트 디바이스(102)의 카메라 시스템은, 클라이언트 디바이스(102)의 전방 및 후방측들 상에 이중 후면 카메라들(예를 들어, 주 카메라뿐만 아니라 깊이 감지(depth-sensing) 카메라), 또는 심지어 삼중, 사중 또는 오중 후면 카메라 구성들을 포함할 수 있다. 이러한 다수의 카메라 시스템들은, 예를 들어, 와이드 카메라, 울트라 와이드 카메라, 망원 사진 카메라, 매크로 카메라, 및 깊이 센서를 포함할 수 있다.

[0133] 위치 컴포넌트들(1034)은 위치 센서 컴포넌트들(예를 들어, GPS 수신기 컴포넌트), 고도 센서 컴포넌트들(예를 들어, 고도계들 또는 고도가 도출될 수 있는 기압을 검출하는 기압계들), 배향 센서 컴포넌트들(예를 들어, 자력계들) 등을 포함한다.

[0134] 통신은 매우 다양한 기술들을 사용하여 구현될 수 있다. I/O 컴포넌트들(1038)은 머신(1000)을 각각의 결합 또는 접속들을 통해 네트워크(1020) 또는 디바이스들(1022)에 결합하도록 동작가능한 통신 컴포넌트들(1036)을 더 포함한다. 예를 들어, 통신 컴포넌트들(1036)은 네트워크 인터페이스 컴포넌트 또는 네트워크(1020)와 인터페이스하기 위한 다른 적절한 디바이스를 포함할 수 있다. 추가 예들에서, 통신 컴포넌트들(1036)은 유선 통신 컴포넌트들, 무선 통신 컴포넌트들, 셀룰러 통신 컴포넌트들, NFC(Near Field Communication) 컴포넌트들, Bluetooth® 컴포넌트들(예를 들어, Bluetooth® Low Energy), Wi-Fi® 컴포넌트들, 및 다른 양상들을 통해 통신을 제공하는 다른 통신 컴포넌트들을 포함할 수 있다. 디바이스(1022)는, 또 다른 머신 또는 임의의 매우 다양한 주변 디바이스(예를 들어, USB를 통해 결합된 주변 디바이스)일 수 있다.

[0135] 더욱이, 통신 컴포넌트들(1036)은 식별자들을 검출할 수 있거나 식별자들을 검출하도록 동작가능한 컴포넌트들을 포함할 수 있다. 예를 들어, 통신 컴포넌트들(1036)은 RFID(Radio Frequency Identification) 태그 판독기 컴포넌트들, NFC 스마트 태그 검출 컴포넌트들, 광학 판독기 컴포넌트들(예를 들어, UPC(Universal Product

Code) 바코드와 같은 1차원 바코드들, QR(Quick Response) 코드와 같은 다차원 바코드들, Aztec 코드, Data Matrix, Dataglyph, MaxiCode, PDF417, Ultra Code, UCC RSS-2D 바코드, 및 다른 광학 코드들을 검출하기 위한 광학 센서), 또는 음향 검출 컴포넌트들(예를 들어, 태깅된 오디오 신호들을 식별하기 위한 마이크로폰들)을 포함할 수 있다. 또한, 인터넷 프로토콜(IP) 지리위치를 통한 위치, Wi-Fi® 신호 삼각측량을 통한 위치, 특정 위치를 나타낼 수 있는 NFC 비컨 신호 검출을 통한 위치 등과 같은 다양한 정보가 통신 컴포넌트들(1036)을 통해 도출될 수 있다.

[0136] 다양한 메모리들(예를 들어, 메인 메모리(1012), 정적 메모리(1014), 및 프로세서들(1002)의 메모리) 및 저장 유닛(1016)은 본 명세서에 설명된 방법론들 또는 기능들 중 어느 하나 이상을 구현하거나 그에 의해 사용되는 명령어들 및 데이터 구조들(예를 들어, 소프트웨어)의 하나 이상의 세트를 저장할 수 있다. 이러한 명령어들(예를 들어, 명령어들(1008))은 프로세서들(1002)에 의해 실행될 때, 다양한 동작들로 하여금 개시된 예들을 구현하게 한다.

[0137] 명령어들(1008)은 송신 매체를 사용하여, 네트워크 인터페이스 디바이스(예를 들어, 통신 컴포넌트들(1036)에 포함된 네트워크 인터페이스 컴포넌트)를 통해 그리고 여러 잘 알려진 전송 프로토콜들(예를 들어, HTTP(hypertext transfer protocol)) 중 어느 하나를 사용하여 네트워크(1020)를 통해 송신 또는 수신될 수 있다. 유사하게, 명령어들(1008)은 디바이스들(1022)에 대한 결합(예를 들어, 피어-투-피어 결합)을 통해 송신 매체를 사용하여 송신되거나 또는 수신될 수 있다.

[0138] 소프트웨어 아키텍처

[0139] 도 11은 본 명세서에 설명된 디바이스들 중 어느 하나 이상에 설치될 수 있는 소프트웨어 아키텍처(1104)를 예시하는 블록도(1100)이다. 소프트웨어 아키텍처(1104)는 프로세서들(1120), 메모리(1126), 및 I/O 컴포넌트들(1138)을 포함하는 머신(1102)과 같은 하드웨어에 의해 지원된다. 이 예에서, 소프트웨어 아키텍처(1104)는 각각의 계층이 특정 기능성을 제공하는 계층들의 스택으로서 개념화될 수 있다. 소프트웨어 아키텍처(1104)는 운영 체제(1112), 라이브러리들(1110), 프레임워크들(1108), 및 애플리케이션(1106)들과 같은 계층들을 포함한다. 동작적으로, 애플리케이션들(1106)은 소프트웨어 스택을 통해 API 호출들(1150)을 기동하고 API 호출들(1150)에 응답하여 메시지들(1152)을 수신한다.

[0140] 운영 체제(1112)는 하드웨어 리소스들을 관리하고 공통 서비스들을 제공한다. 운영 체제(1112)는, 예를 들어, 커널(1114), 서비스들(1116), 및 드라이버들(1122)을 포함한다. 커널(1114)은 하드웨어와 다른 소프트웨어 계층들 사이의 추상화 계층으로서 작용한다. 예를 들어, 커널(1114)은, 다른 기능 중에서, 메모리 관리, 프로세서 관리(예를 들어, 스케줄링), 컴포넌트 관리, 네트워크화, 및 보안 설정을 제공한다. 서비스들(1116)은 다른 소프트웨어 계층들을 위한 다른 공통 서비스들을 제공할 수 있다. 드라이버들(1122)은 기본 하드웨어를 제어하거나 그와 인터페이싱하는 것을 담당한다. 예를 들어, 드라이버들(1122)은 디스플레이 드라이버들, 카메라 드라이버들, BLUETOOTH® 또는 BLUETOOTH® Low Energy 드라이버들, 플래시 메모리 드라이버들, 직렬 통신 드라이버들(예를 들면, USB 드라이버들), Wi-Fi® 드라이버들, 오디오 드라이버들, 전력 관리 드라이버들 등을 포함할 수 있다.

[0141] 라이브러리들(1110)은 애플리케이션들(1106)에 의해 사용되는 공통 로우-레벨 인프라스트럭처를 제공한다. 라이브러리들(1110)은, 메모리 할당 기능들, 문자열 조작 기능들, 수학 기능들 등과 같은 기능들 제공할 수 있는 시스템 라이브러리들(1118)(예를 들어, C 표준 라이브러리)를 포함할 수 있다. 또한, 라이브러리들(1110)은, 미디어 라이브러리들(예를 들어, MPEG4(Moving Picture Experts Group-4), H.264 또는 AVC(Advanced Video Coding), MP3(Moving Picture Experts Group Layer-3), AAC(Advanced Audio Coding), AMR(Adaptive Multi-Rate) 오디오 코덱, JPG 또는 JPEG(Joint Photographic Experts Group), 또는 PNG(Portable Network Graphics)와 같은 다양한 미디어 포맷들의 제시 및 조작을 지원하는 라이브러리들), 그래픽 라이브러리들(예를 들어, 디스플레이 상의 그래픽 콘텐츠에서 2D 및 3D으로 렌더링하기 위해 사용되는 OpenGL 프레임워크), 데이터베이스 라이브러리들(예를 들어, 다양한 관계형 데이터베이스 기능들을 제공하는 SQLite), 웹 라이브러리들(예를 들어, 웹 브라우징 기능성을 제공하는 WebKit) 등과 같은 API 라이브러리들(1124)을 포함할 수 있다. 라이브러리들(1110)은 또한, 많은 다른 API를 애플리케이션(1106)에 제공하는 매우 다양한, 다른 라이브러리(1128)를 포함할 수 있다.

[0142] 프레임워크들(1108)은 애플리케이션들(1106)에 의해 사용되는 공통 고레벨 인프라스트럭처(high-level infrastructure)를 제공한다. 예를 들어, 프레임워크들(1108)은, 다양한 그래픽 사용자 인터페이스 기능, 하이-레벨 리소스 관리, 및 하이-레벨 위치 서비스를 제공한다. 프레임워크들(1108)은 애플리케이션(1106)에 의해

이용될 수 있는 광범위한 스펙트럼의 다른 API들을 제공할 수 있고, 그 중 일부는 특정한 운영 체제 또는 플랫폼 특정적일 수 있다.

[0143] 예에서, 애플리케이션들(1106)은 홈 애플리케이션(1136), 연락처 애플리케이션(1130), 브라우저 애플리케이션(1132), 북 리더 애플리케이션(1134), 위치 애플리케이션(1142), 미디어 애플리케이션(1144), 메시징 애플리케이션(1146), 게임 애플리케이션(1148), 및 외부 애플리케이션(1140)과 같은 광범위한 다른 애플리케이션들을 포함할 수 있다. 애플리케이션들(1106)은 프로그램들에 정의된 함수들을 실행하는 프로그램들이다. 객체 지향(object-oriented) 프로그램 언어들(예를 들어, Objective-C, Java, 또는 C++) 또는 절차적 프로그램 언어들(procedural programming language)(예를 들어, C 또는 어셈블리 언어)과 같은, 다양한 방식으로 구조화된, 애플리케이션들(1106) 중 하나 이상을 생성하기 위해 다양한 프로그램 언어들이 이용될 수 있다. 특정 예에서, 외부 애플리케이션(1140)(예를 들어, 특정 플랫폼의 벤더 이외의 엔티티에 의해 ANDROID™ 또는 IOS™ SDK를 사용하여 개발되는 애플리케이션)은 IOS™, ANDROID™, WINDOWS® Phone, 또는 다른 모바일 운영 체제와 같은 모바일 운영 체제 상에서 실행되는 모바일 소프트웨어일 수 있다. 이 예에서, 외부 애플리케이션(1140)은 본 명세서에 설명된 기능성을 용이하게 하기 위해 운영 체제(1112)에 의해 제공되는 API 호출들(1150)을 기동할 수 있다.

[0144] 용어

[0145] "캐리어 신호(carrier signal)"는 머신에 의한 실행을 위한 명령어들을 저장, 인코딩, 또는 운반할 수 있는 임의의 무형 매체를 지칭하고, 이러한 명령어들의 통신을 용이하게 하기 위한 디지털 또는 아날로그 통신 신호들 또는 다른 무형 매체를 포함한다. 명령어들은 네트워크 인터페이스 디바이스를 통해 송신 매체를 사용하여 네트워크를 통해 송신되거나 수신될 수 있다.

[0146] "클라이언트 디바이스"는 하나 이상의 서버 시스템 또는 다른 클라이언트 디바이스들로부터 리소스들을 획득하기 위해 통신 네트워크에 인터페이스하는 임의의 머신을 지칭한다. 클라이언트 디바이스는, 이동 전화, 데스크톱 컴퓨터, 랩톱, PDA, 스마트폰, 태블릿, 울트라북, 넷북, 랩톱, 멀티-프로세서 시스템, 마이크로프로세서-기반 또는 프로그램가능한 소비자 가전 제품, 게임 콘솔, 셋톱 박스, 또는 사용자가 네트워크에 액세스하기 위해 사용할 수 있는 임의의 다른 통신 디바이스일 수 있지만, 이에 제한되지 않는다.

[0147] "통신 네트워크"는 애드혹 네트워크, 인트라넷, 엑스트라넷, VPN(virtual private network), LAN(local area network), WLAN(wireless LAN), WAN(wide area network), WWAN(wireless WAN), MAN(metropolitan area network), 인터넷, 인터넷의 일부, PSTN(Public Switched Telephone Network)의 일부, POTS(plain old telephone service) 네트워크, 셀룰러 전화 네트워크, 무선 네트워크, Wi-Fi® 네트워크, 다른 타입의 네트워크, 또는 2개 이상의 이러한 네트워크의 조합일 수 있는 네트워크의 하나 이상의 부분을 지칭한다. 예를 들어, 네트워크 또는 네트워크의 일부는 무선 또는 셀룰러 네트워크를 포함할 수 있고, 결합은 CDMA(Code Division Multiple Access) 접속, GSM(Global System for Mobile communications) 접속, 또는 다른 타입들의 셀룰러 또는 무선 결합일 수 있다. 이 예에서, 결합은 1xRTT(Single Carrier Radio Transmission Technology), EVDO(Evolution-Data Optimized) 기술, GPRS(General Packet Radio Service) 기술, EDGE(Enhanced Data rates for GSM Evolution) 기술, 3G를 포함하는 3GPP(third Generation Partnership Project), 4G(fourth generation wireless) 네트워크들, UMTS(Universal Mobile Telecommunications System), HSPA(High Speed Packet Access), WiMAX(Worldwide Interoperability for Microwave Access), LTE(Long Term Evolution) 표준, 다양한 표준 설정 조직들에 의해 정의된 다른 것들, 다른 장거리 프로토콜들, 또는 다른 데이터 전송 기술과 같은, 다양한 타입의 데이터 전송 기술들 중 임의의 것을 구현할 수 있다.

[0148] "컴포넌트(Component)"는 기능 또는 서브루틴 호출들, 분기 포인트들, API들, 또는 특정 처리 또는 제어 기능들의 파티셔닝 또는 모듈화를 제공하는 다른 기술들에 의해 정의되는 경계들을 갖는 디바이스, 물리적 엔티티, 또는 로직을 지칭한다. 컴포넌트들은 그들의 인터페이스들을 통해 다른 컴포넌트들과 조합되어 머신 프로세스를 수행할 수 있다. 컴포넌트는 다른 컴포넌트들 및 일반적으로 관련된 기능들 중 특정 기능을 수행하는 프로그램의 일부와 함께 사용하도록 설계되는 패키징된 기능적 하드웨어 유닛일 수 있다.

[0149] 컴포넌트들은 소프트웨어 컴포넌트들(예를 들어, 머신 판독가능 매체 상에 구체화된 코드) 또는 하드웨어 컴포넌트들 중 하나를 구성할 수 있다. "하드웨어 컴포넌트"는 특정 동작들을 수행할 수 있는 유형적 유닛(tangible unit)이고, 특정 물리적 방식으로 구성되거나 배열될 수 있다. 다양한 예들에서, 하나 이상의 컴퓨터 시스템(예를 들어, 독립형 컴퓨터 시스템, 클라이언트 컴퓨터 시스템, 또는 서버 컴퓨터 시스템) 또는 컴퓨

터 시스템의 하나 이상의 하드웨어 컴포넌트(예를 들어, 프로세서 또는 프로세서들의 그룹)는 본 명세서에 설명된 바와 같은 특정 동작들을 수행하도록 동작하는 하드웨어 컴포넌트로서 소프트웨어(예를 들어, 애플리케이션 또는 애플리케이션 부분)에 의해 구성될 수 있다.

[0150] 하드웨어 컴포넌트는 또한, 기계적으로, 전자적으로, 또는 이들의 임의의 적합한 조합으로 구현될 수 있다. 예를 들어, 하드웨어 컴포넌트는 특정 동작들을 수행하도록 영구적으로 구성된 전용 회로 또는 로직을 포함할 수 있다. 하드웨어 컴포넌트는 FPGA(field-programmable gate array) 또는 ASIC(application specific integrated circuit)와 같은 특수 목적 프로세서일 수 있다. 하드웨어 컴포넌트는 또한, 특정 동작들을 수행하도록 소프트웨어에 의해 일시적으로 구성된 프로그램가능 로직 또는 회로부를 포함할 수 있다. 예를 들어, 하드웨어 컴포넌트는 범용 프로세서 또는 다른 프로그램가능 프로세서에 의해 실행되는 소프트웨어를 포함할 수 있다. 이러한 소프트웨어에 의해 구성되고 나면, 하드웨어 컴포넌트들은 구성된 기능들을 수행하도록 독자적으로 맞춤화된(uniquely tailored) 특정 머신들(또는 머신의 특정 컴포넌트들)이 되고 더 이상 범용 프로세서들이 아니다. 하드웨어 컴포넌트를 기계적으로, 전용의 영구적으로 구성된 회로에, 또는 일시적으로 구성된 회로(예를 들어, 소프트웨어에 의해 구성됨)에 구현하기로 하는 결정은 비용 및 시간 고려사항들에 의해 주도될 수 있다는 것을 이해할 것이다. 따라서, "하드웨어 컴포넌트"(또는 "하드웨어 구현 컴포넌트(hardware-implemented component)")라는 문구는, 유형 엔티티, 즉, 특정 방식으로 동작하거나 본 명세서에 설명된 특정 동작들을 수행하도록 물리적으로 구성되거나, 영구적으로 구성되거나(예를 들어, 하드와이어드) 또는 일시적으로 구성되는(예를 들어, 프로그램되는) 엔티티를 포괄하는 것으로 이해해야 한다.

[0151] 하드웨어 컴포넌트들이 일시적으로 구성되는(예를 들어, 프로그램되는) 예들을 고려하면, 하드웨어 컴포넌트들 각각은 임의의 한 시점에서 구성되거나 인스턴스화될 필요는 없다. 예를 들어, 하드웨어 컴포넌트가 소프트웨어에 의해 특수 목적 프로세서가 되도록 구성된 범용 프로세서를 포함하는 경우, 범용 프로세서는 상이한 시간들에서 (예를 들어, 상이한 하드웨어 컴포넌트들을 포함하는) 각각 상이한 특수 목적 프로세서들로서 구성될 수 있다. 따라서 소프트웨어는, 예를 들어, 특정 프로세서 또는 프로세서들을 구성하여 하나의 시간 인스턴스에서 특정 하드웨어 컴포넌트를 구성하고 상이한 시간 인스턴스에서 상이한 하드웨어 컴포넌트를 구성한다.

[0152] 하드웨어 컴포넌트들은 다른 하드웨어 컴포넌트들에 정보를 제공하고 다른 하드웨어 컴포넌트들로부터 정보를 수신할 수 있다. 그에 따라, 설명된 하드웨어 컴포넌트들은 통신가능하게 결합되어 있는 것으로 간주될 수 있다. 다수의 하드웨어 컴포넌트들이 동시에 존재하는 경우에, (예를 들어, 적절한 회로들 및 버스들을 통한) 하드웨어 컴포넌트들 중 2개 이상 사이의 신호 송신을 통해 통신이 달성될 수 있다. 다수의 하드웨어 컴포넌트가 상이한 시간들에서 구성되거나 인스턴스화되는 예들에서, 이러한 하드웨어 컴포넌트들 사이의 통신은, 예를 들어, 다수의 하드웨어 컴포넌트들이 액세스할 수 있는 메모리 구조들에서의 정보의 저장 및 검색을 통해 달성될 수 있다. 예를 들어, 하나의 하드웨어 컴포넌트는 동작을 수행하고, 그에 통신가능하게 결합되는 메모리 디바이스에 해당 동작의 출력을 저장할 수 있다. 이어서, 추가의 하드웨어 컴포넌트는, 나중에, 메모리 디바이스에 액세스하여 저장된 출력을 검색 및 처리할 수 있다. 하드웨어 컴포넌트들은 또한 입력 또는 출력 디바이스들과의 통신들을 개시할 수 있고, 리소스(예를 들어, 정보의 컬렉션) 상에서 동작할 수 있다.

[0153] 본 명세서에 설명된 예시적인 방법들의 다양한 동작들은, 관련 동작들을 수행하도록 일시적으로 구성되거나(예컨대, 소프트웨어에 의함) 또는 영구적으로 구성되는 하나 이상의 프로세서들에 의해 적어도 부분적으로 수행될 수 있다. 일시적으로 구성되든 또는 영구적으로 구성되든 간에, 이러한 프로세서들은 본 명세서에 설명되는 하나 이상의 동작 또는 기능을 수행하도록 동작하는 프로세서 구현 컴포넌트들을 구성할 수 있다. 본 명세서에서 사용되는 바와 같이, "프로세서 구현 컴포넌트"는 하나 이상의 프로세서들을 사용하여 구현된 하드웨어 컴포넌트를 지칭한다. 유사하게, 본 명세서에 설명된 방법들은 특정 프로세서 또는 프로세서들을 하드웨어의 일 예로 하여 적어도 부분적으로 프로세서 구현될 수 있다. 예를 들어, 방법의 동작들 중 적어도 일부가 하나 이상의 프로세서(1002) 또는 프로세서 구현 컴포넌트에 의해 수행될 수 있다. 더욱이, 하나 이상의 프로세서들은 "클라우드 컴퓨팅" 환경에서 또는 "SaaS(software as a service)"로서 관련 동작들의 수행을 지원하도록 또한 동작할 수 있다. 예를 들어, 동작들 중 적어도 일부는 (프로세서들을 포함하는 머신들의 예들로서) 컴퓨터들의 그룹에 의해 수행될 수 있고, 이러한 동작들은 네트워크(예를 들어, 인터넷)를 통해 그리고 하나 이상의 적절한 인터페이스(예를 들어, API)를 통해 액세스가능하다. 동작들 중 특정한 동작의 수행은 단일 머신 내에 상주할 뿐만 아니라 다수의 머신들에 걸쳐 배치되는 프로세서들 간에 분산될 수 있다. 일부 예들에서, 프로세서들 또는 프로세서 구현 컴포넌트들은 단일의 지리적 위치에(예컨대, 가정 환경, 사무실 환경, 또는 서버 팜 내에) 위치될 수 있다. 다른 예들에서, 프로세서들 또는 프로세서 구현 컴포넌트들은 다수의 지리적 위치에 걸쳐 분산될 수 있다.

[0154] "컴퓨터 판독가능 저장 매체"는 머신 저장 매체 및 전송 매체 둘 다를 지칭한다. 따라서, 용어들은 저장 디바이스들/매체 및 반송파들/변조된 데이터 신호들 둘 모두를 포함한다. "머신 판독가능 매체", "컴퓨터 판독가능 매체" 및 "디바이스 판독가능 매체"라는 용어들은 동일한 것을 의미하고, 본 개시내용에서 상호교환가능하게 사용될 수 있다.

[0155] "단기적 메시지"는 시간-제한된 지속기간 동안 액세스가능한 메시지를 지칭한다. 단기적 메시지는 텍스트, 이미지, 비디오 등일 수 있다. 단기적 메시지에 대한 액세스 시간은 메시지 발신자에 의해 설정될 수 있다. 대안적으로, 액세스 시간은 디폴트 설정이거나 수신인에 의해 지정된 설정일 수 있다. 설정 기술에 관계없이, 메시지는 일시적이다.

[0156] "머신 저장 매체(machine storage medium)"는 실행가능한 명령어들, 루틴들 및 데이터를 저장하는 단일 또는 다수의 저장 디바이스들 및 매체들(예를 들어, 중앙집중형 또는 분산형 데이터베이스, 및 연관된 캐시들 및 서버들)을 지칭한다. 따라서, 이 용어는 프로세서들 내부 또는 외부의 메모리를 포함하여, 솔리드-스테이트 메모리들, 및 광학 및 자기 매체를 포함하지만 이에 제한되지 않는 것으로 취해질 것이다. 머신 저장 매체들, 컴퓨터 저장 매체들 및 디바이스 저장 매체들의 특정 예들은 예로서 반도체 메모리 디바이스들, 예를 들어, EPROM(erasable programmable read-only memory), EEPROM(electrically erasable programmable read-only memory), FPGA, 및 플래시 메모리 디바이스들을 포함하는 비휘발성 메모리; 내부 하드 디스크 및 착탈식 디스크와 같은 자기 디스크; 광자기 디스크들; 그리고 CD-ROM 및 DVD-ROM 디스크들을 포함한다. "머신 저장 매체", "디바이스 저장 매체", "컴퓨터 저장 매체"라는 용어들은 동일한 것을 의미하고, 본 개시내용에서 상호교환가능하게 사용될 수 있다. "머신-저장 매체(machine-storage media)", "컴퓨터-저장 매체(computer-storage media)", 및 "디바이스-저장 매체(device-storage media)"라는 용어들은 반송파들, 변조된 데이터 신호들, 및 다른 이러한 매체를 구체적으로 배제하고, 이들 중 적어도 일부는 "신호 매체(signal medium)"라는 용어에 커버된다.

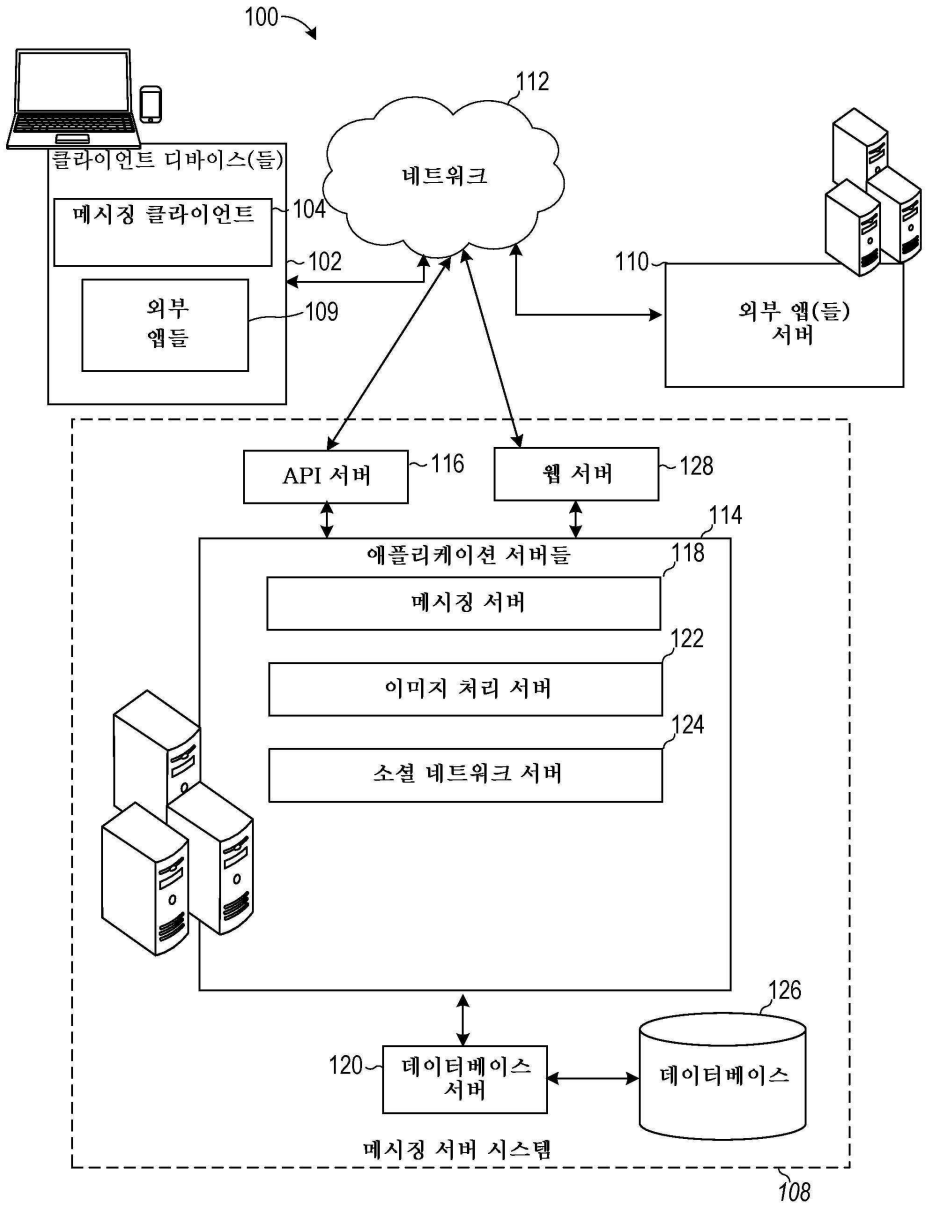
[0157] "비일시적 컴퓨터 판독가능 저장 매체(non-transitory computer-readable storage medium)"는 머신에 의한 실행을 위한 명령어들을 저장, 인코딩, 또는 운반할 수 있는 유형의 매체를 지칭한다.

[0158] "신호 매체"는 머신에 의한 실행을 위한 명령어들을 저장, 인코딩, 또는 운반할 수 있는 임의의 무형 매체를 지칭하고, 소프트웨어 또는 데이터의 통신을 용이하게 하기 위한 디지털 또는 아날로그 통신 신호들 또는 다른 무형 매체를 포함한다. 용어 "신호 매체"는 임의의 형태의 변조된 데이터 신호, 반송파 등을 포함하는 것으로 고려되어야 한다. "변조된 데이터 신호(modulated data signal)"라는 용어는 신호에 정보를 인코딩하는 것과 같은 이러한 문제에서 설정 또는 변경된 그 특성들 중 하나 이상을 갖는 신호를 의미한다. "송신 매체(transmission medium)", 및 "신호 매체(signal medium)"이라는 용어들은 동일한 것을 의미하고 본 개시내용에서 상호교환가능하게 사용될 수 있다.

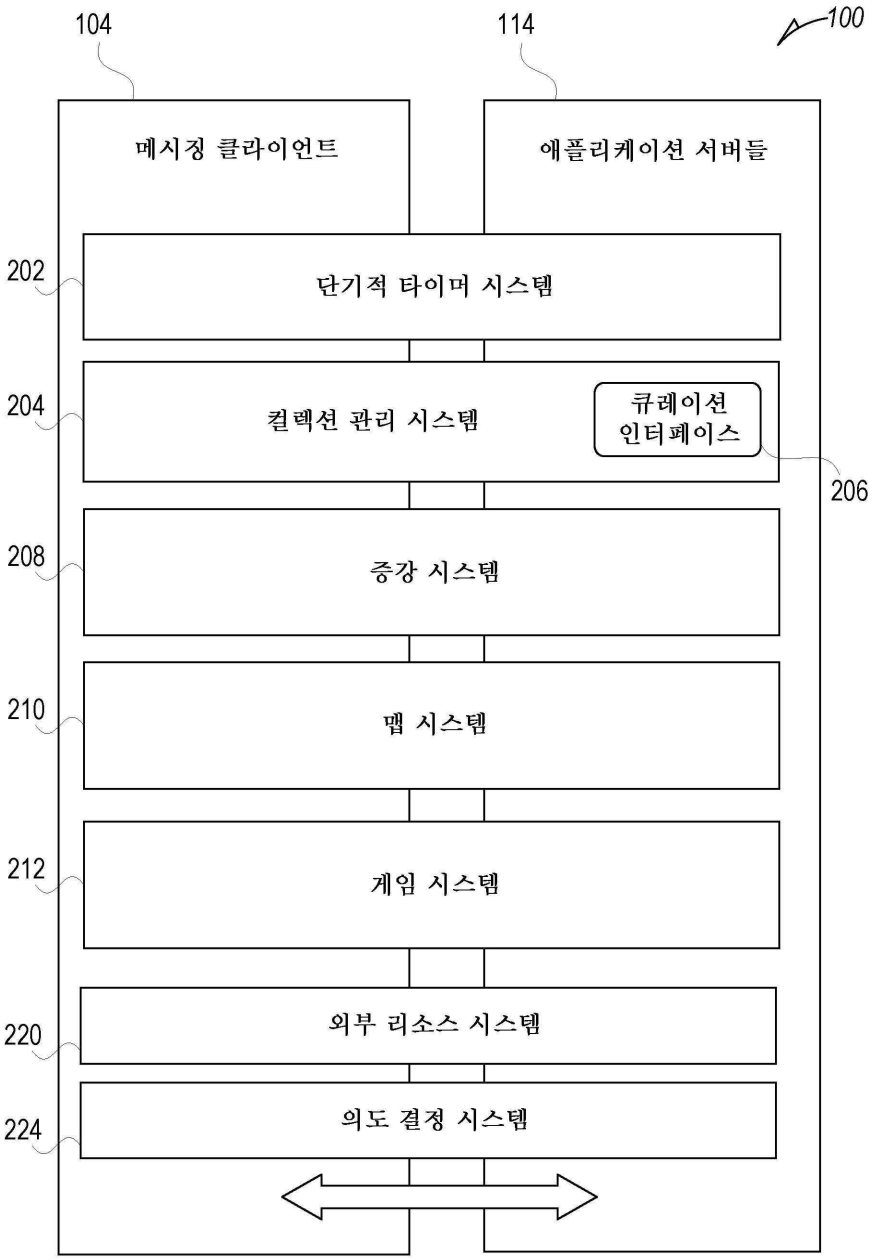
[0159] 본 개시내용의 범위를 벗어나지 않으면서 개시된 예들에 대해 변경들 및 수정들이 이루어질 수 있다. 이들 및 다른 변경들 또는 수정들은 다음의 청구항들에서 표현된 바와 같은, 본 개시내용의 범위 내에 포함되는 것을 의도한다.

도면

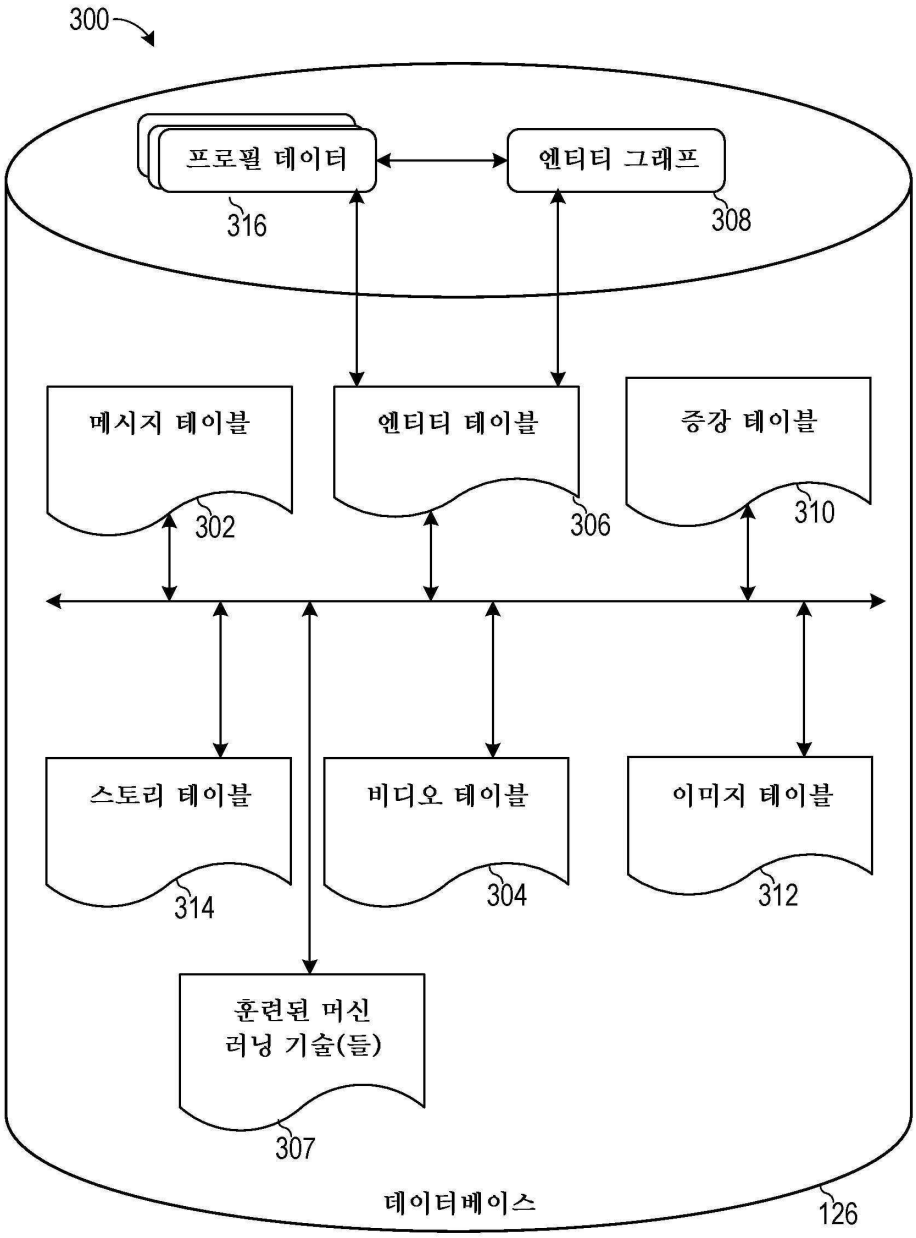
도면1


100
112
네트워크
클라이언트 디바이스(들)
메시징 클라이언트
104
102
외부
앱들
109
110
외부 앱(들)
서버
API 서버
116
웹 서버
128
114
애플리케이션 서버들
118
메시징 서버
122
이미지 처리 서버
124
소셜 네트워크 서버
126
120
데이터베이스
서버
데이터베이스
메시징 서버 시스템
108

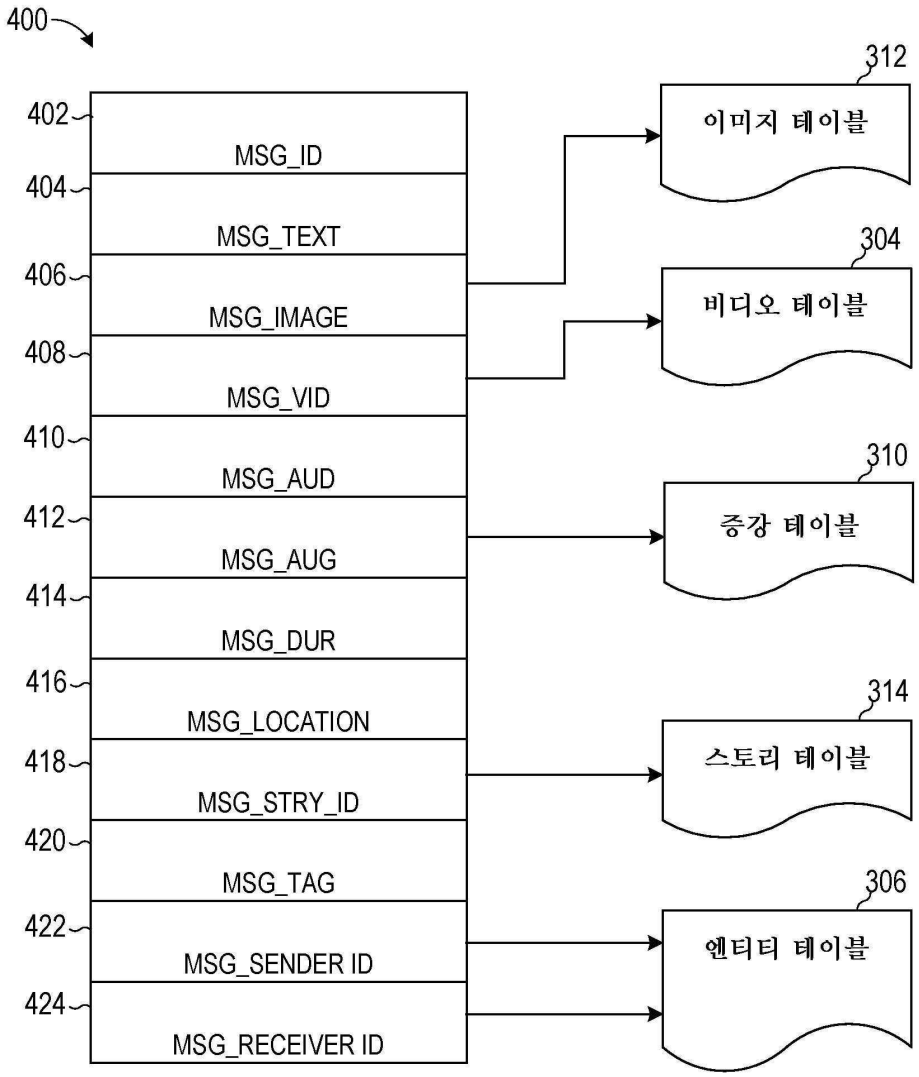
도면2


104
114
100
메시징 클라이언트
애플리케이션 서버들
202
단기적 타이머 시스템
204
컬렉션 관리 시스템
큐레이션 인터페이스
206
208
증강 시스템
210
맵 시스템
212
게임 시스템
220
외부 리소스 시스템
224
의도 결정 시스템

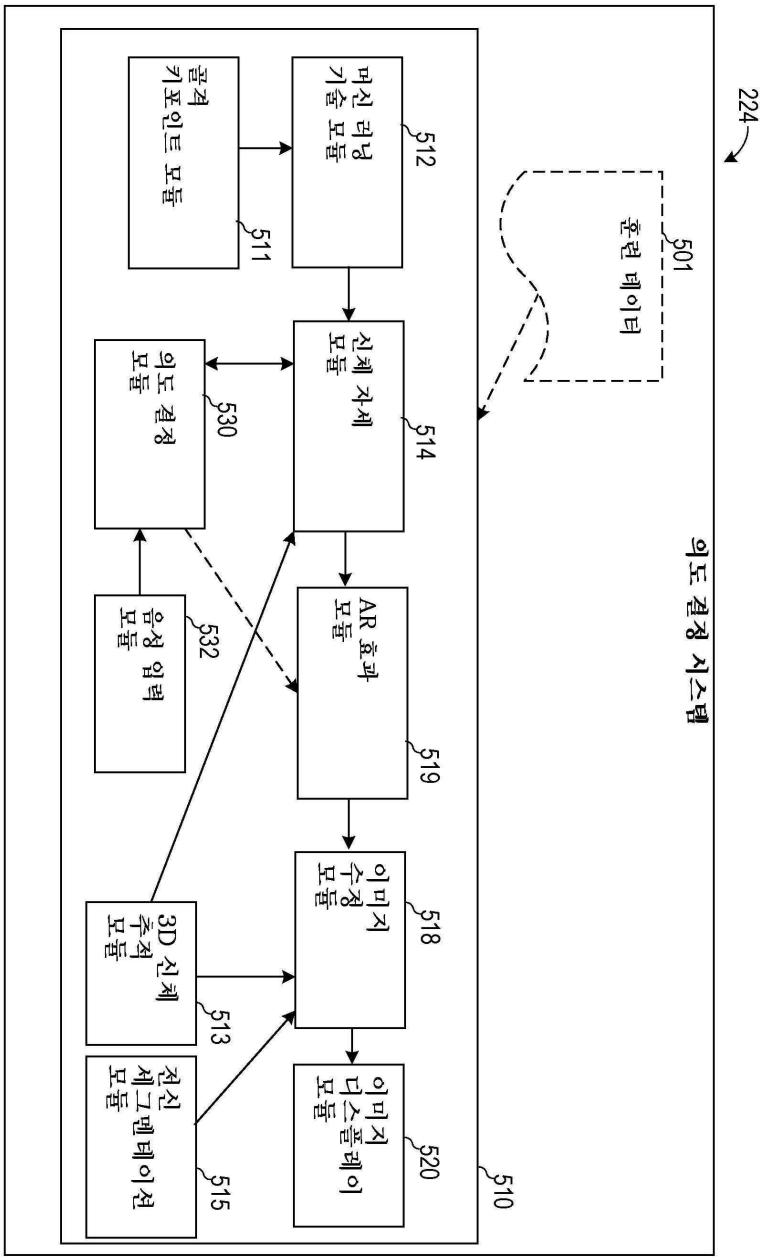
도면3


300
프로필 데이터
316
엔티티 그래프
308
메시지 테이블
302
엔티티 테이블
306
증강 테이블
310
스토리 테이블
314
비디오 테이블
304
이미지 테이블
312
훈련된 머신
러닝 기술(들)
307
데이터베이스
126

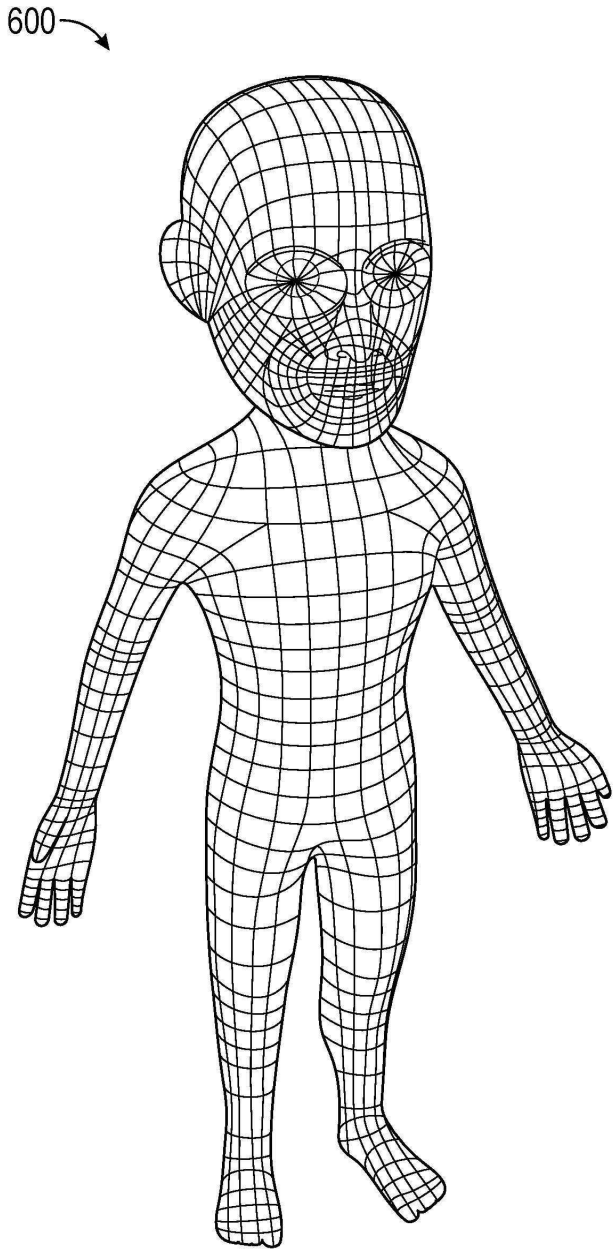
도면4


400
402
MSG_ID
404
MSG_TEXT
406
MSG_IMAGE
408
MSG_VID
410
MSG_AUD
412
MSG_AUG
414
MSG_DUR
416
MSG_LOCATION
418
MSG_STRY_ID
420
MSG_TAG
422
MSG_SENDER ID
424
MSG_RECEIVER ID
312
이미지 테이블
304
비디오 테이블
310
증강 테이블
314
스토리 테이블
306
엔티티 테이블

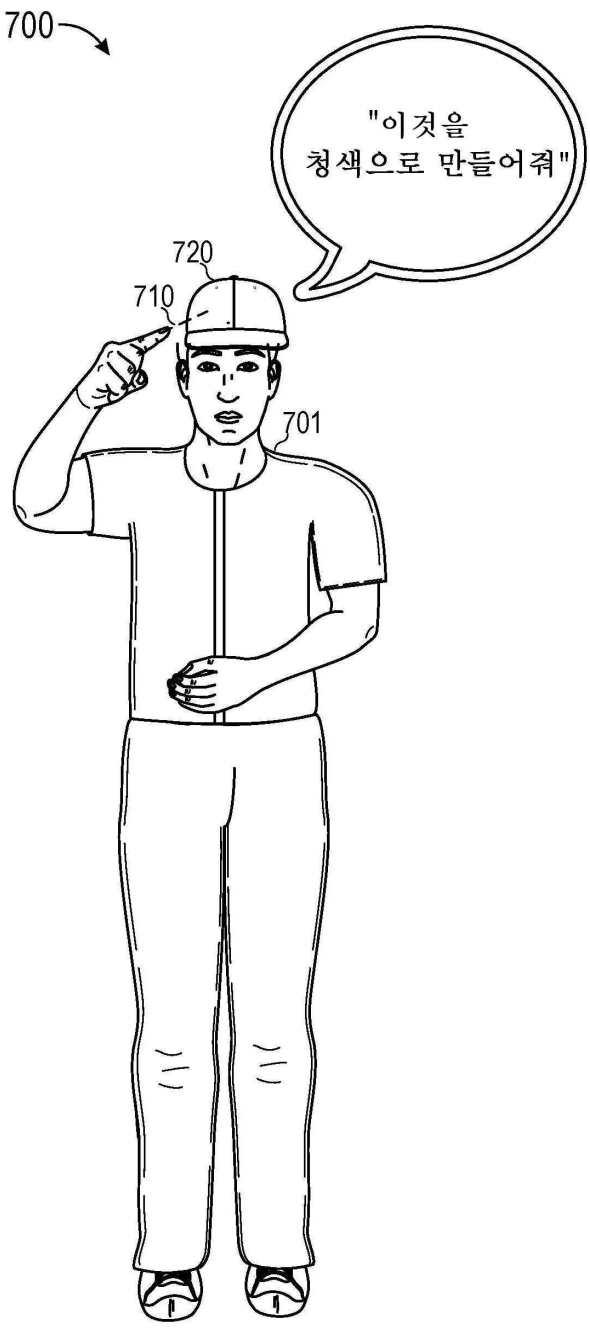
도면5


224
의도 결정 시스템
훈련 데이터
501
머신 러닝 기술 모듈
512
골격 키포인트 모듈
511
신체 자세 모듈
514
의도 결정 모듈
530
음성 입력 모듈
532
AR 효과 모듈
519
이미지 수정 모듈
518
3D 신체 추적 모듈
513
전신 세그멘테이션 모듈
515
이미지 디스플레이 모듈
520
510

도면6



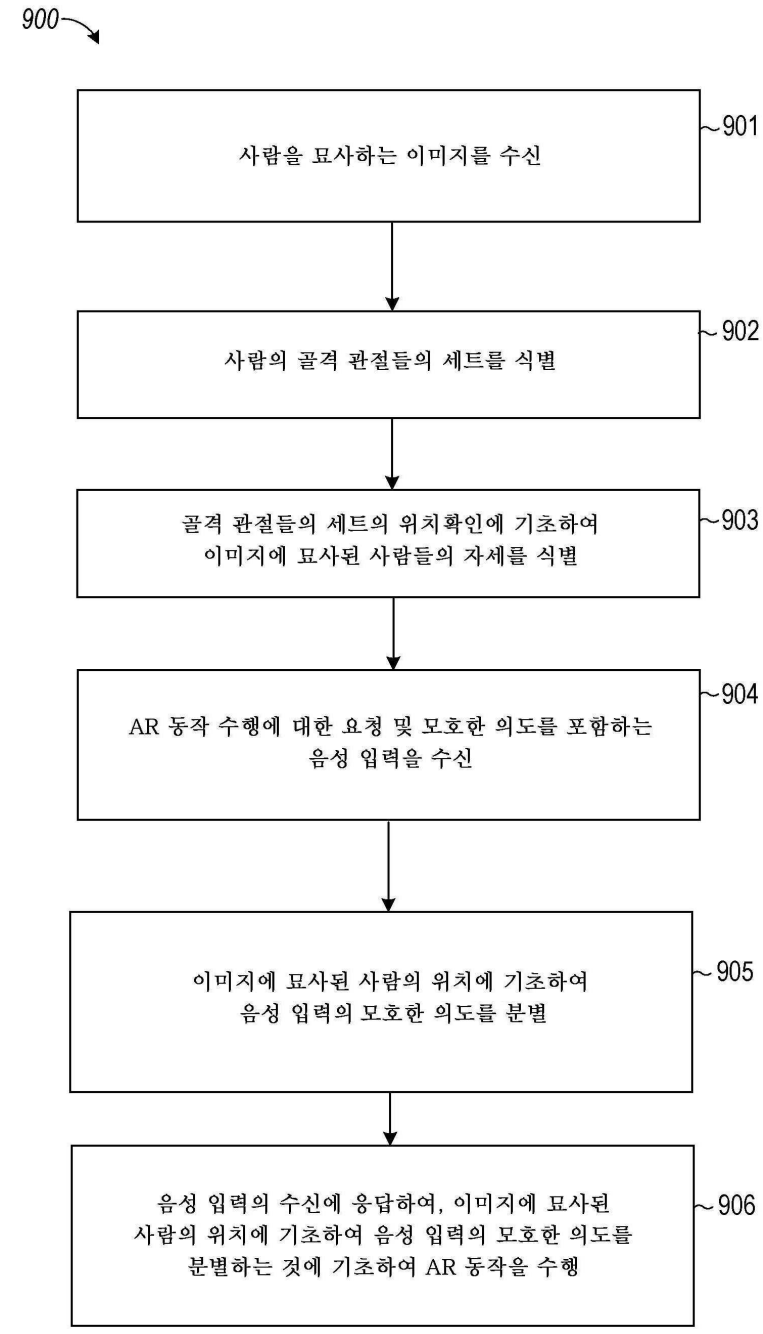
도면7



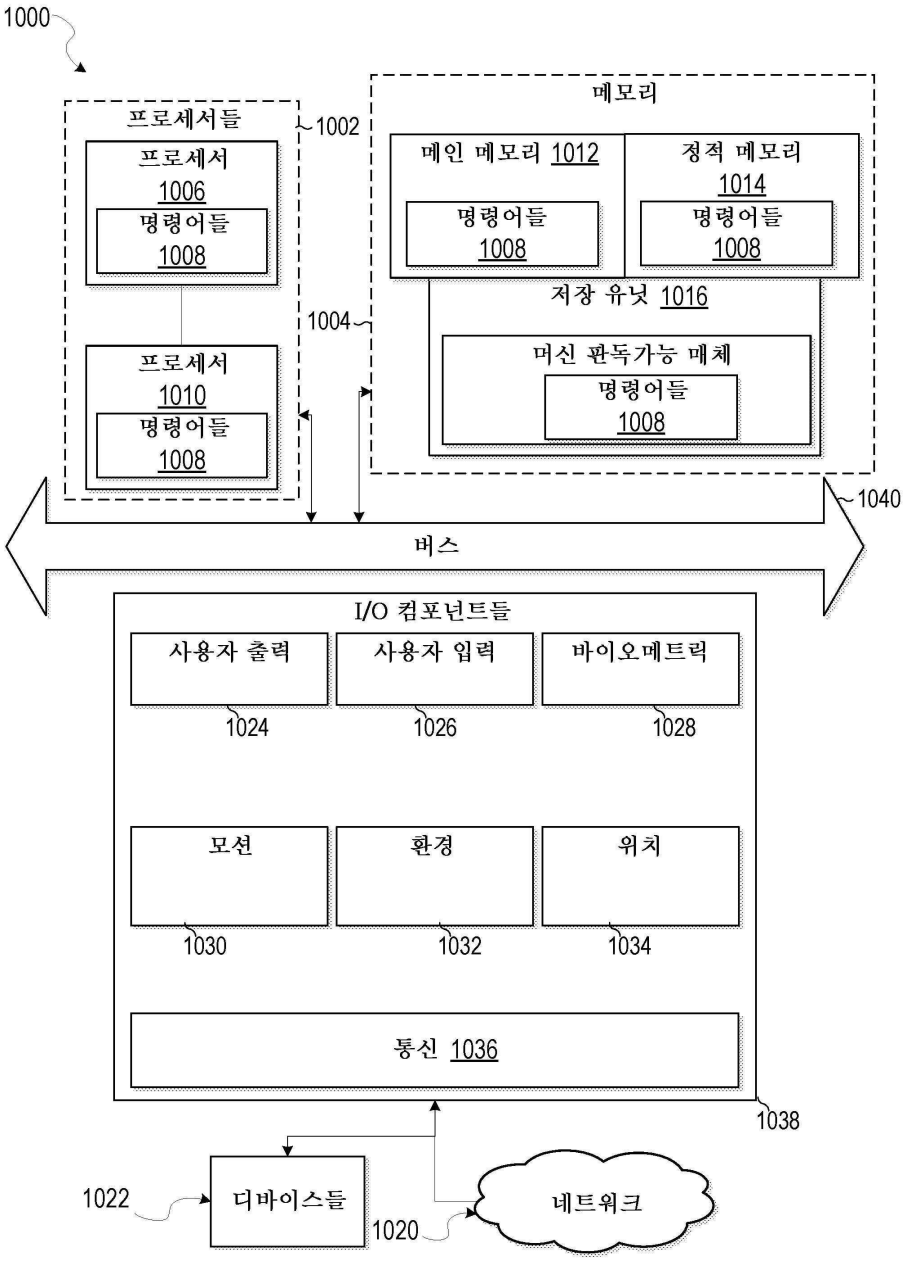
도면8



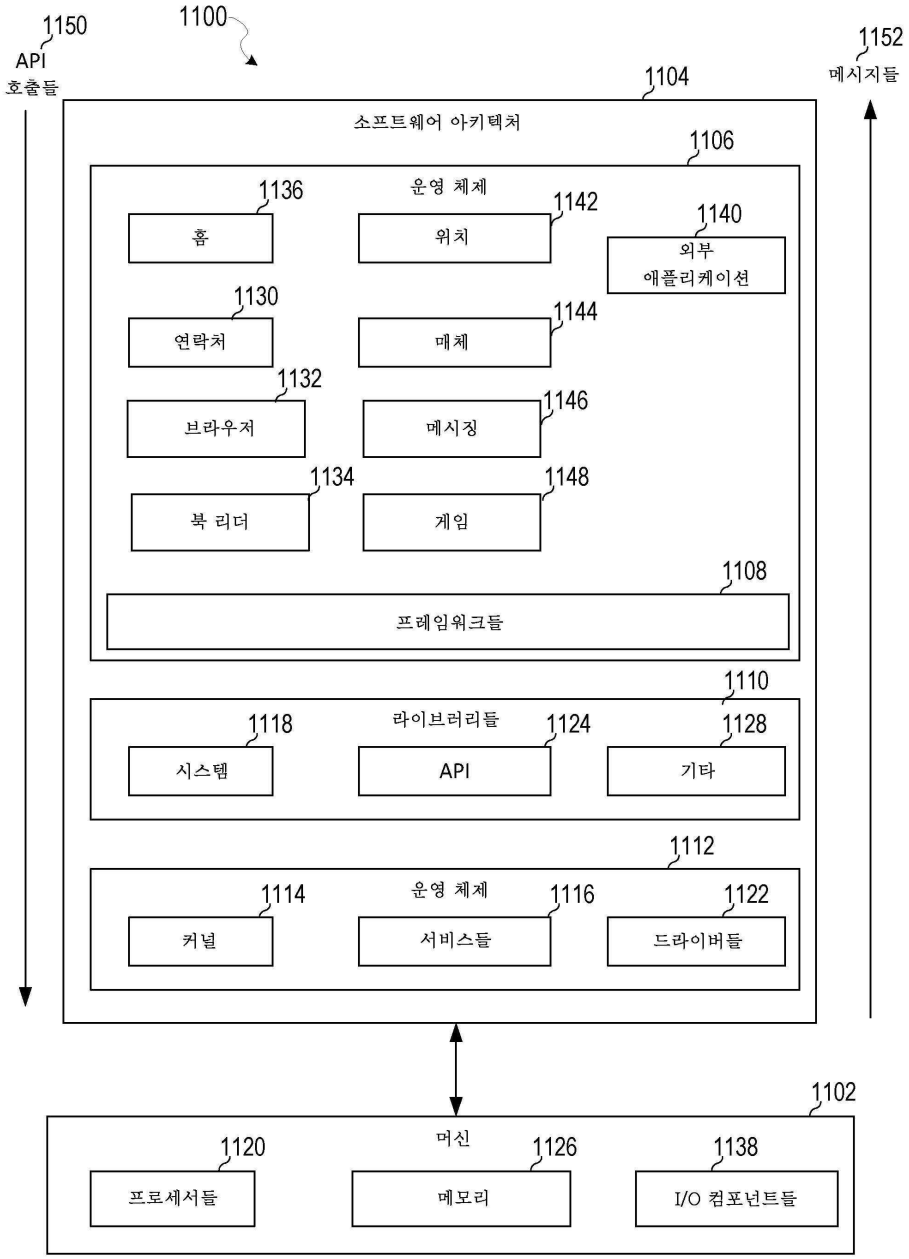
도면9


900
사람을 묘사하는 이미지를 수신
901
사람의 골격 관절들의 세트를 식별
902
골격 관절들의 세트의 위치확인에 기초하여
이미지에 묘사된 사람들의 자세를 식별
903
AR 동작 수행에 대한 요청 및 모호한 의도를 포함하는
음성 입력을 수신
904
이미지에 묘사된 사람의 위치에 기초하여
음성 입력의 모호한 의도를 분별
905
음성 입력의 수신에 응답하여, 이미지에 묘사된
사람의 위치에 기초하여 음성 입력의 모호한 의도를
분별하는 것에 기초하여 AR 동작을 수행
906

도면10


1000
프로세서들
1002
프로세서
1006
명령어들
1008
프로세서
1010
명령어들
1008
메모리
메인 메모리 1012
명령어들
1008
정적 메모리
1014
명령어들
1008
저장 유닛 1016
머신 판독가능 매체
명령어들
1008
1004
1040
버스
I/O 컴포넌트들
사용자 출력
사용자 입력
바이오메트릭
1024
1026
1028
모션
환경
위치
1030
1032
1034
통신 1036
1038
1022
디바이스들
1020
네트워크

도면11


1150
API
호출들
1100
1152
메시지들
1104
소프트웨어 아키텍처
1106
운영 체제
1136
홈
1142
위치
1140
외부
애플리케이션
1130
연락처
1144
매체
1132
브라우저
1146
메시징
1134
북 리더
1148
게임
1108
프레임워크들
1110
라이브러리들
1118
시스템
1124
API
1128
기타
1112
운영 체제
1114
커널
1116
서비스들
1122
드라이버들
1102
머신
1120
프로세서들
1126
메모리
1138
I/O 컴포넌트들