



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년03월07일
(11) 등록번호 10-2506959
(24) 등록일자 2023년03월02일

(51) 국제특허분류(Int. Cl.)
G06T 7/593 (2017.01) G06T 7/00 (2017.01)
G06T 7/30 (2017.01)
(52) CPC특허분류
G06T 7/593 (2017.01)
G06T 7/30 (2017.01)
(21) 출원번호 10-2020-7036300
(22) 출원일자(국제) 2019년05월16일
심사청구일자 2022년05월13일
(85) 번역문제출일자 2020년12월16일
(65) 공개번호 10-2020-0146040
(43) 공개일자 2020년12월31일
(86) 국제출원번호 PCT/US2019/032616
(87) 국제공개번호 WO 2019/222467
국제공개일자 2019년11월21일
(30) 우선권주장
62/673,045 2018년05월17일 미국(US)
(56) 선행기술조사문헌
Tinghui Zhou 등, Unsupervised Learning of
Depth and Ego-Motion from Video,
arXiv:1704.07813v2(2017.08.01.)*
Benjamin Ummenhofer 등, DeMoN: Depth and
Motion Network for Learning Monocular Stereo,
arXiv:1612.02401v2(2017.04.11.)
W02018046964 A1
KR1020170120639 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
나이앤틱, 인크.
미국, 캘리포니아주 94111, 샌프란시스코, 스위트
200, 원 페리 빌딩
(72) 발명자
고다드, 클레멘트
미국, 캘리포니아주 94111, 샌프란시스코, 스위트
200, 원 페리 빌딩, 나이앤틱, 인크. 사내
맥 아오다, 오이신
미국, 캘리포니아주 94111, 샌프란시스코, 스위트
200, 원 페리 빌딩, 나이앤틱, 인크. 사내
(뒷면에 계속)
(74) 대리인
김영철, 김 순 영

전체 청구항 수 : 총 20 항

심사관 : 이재원

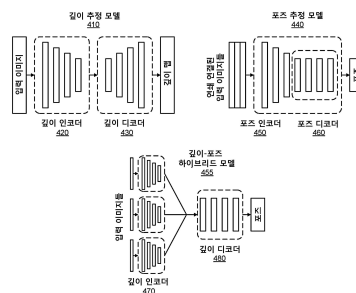
(54) 발명의 명칭 깊이 추정 시스템의 자가 감독 훈련

(57) 요약

깊이 추정 모델을 훈련하는 방법 및 상기 깊이 추정 모델을 사용하는 방법을 기술한다. 이미지들을 획득하여 깊이 모델에 입력하여서 깊이 모델의 파라미터들에 기초하여 복수의 이미지들 각각에 대한 깊이 맵을 추출한다. 본 방법은 이미지들을 포즈 디코더에 입력하여 각 이미지에 대한 포즈를 추출하는 단계를 포함한다. 본 방법은 각

(뒷면에 계속)

대표도 - 도4



이미지에 대한 포즈와 깊이 맵에 기초하여 복수의 합성 프레임을 생성하는 단계를 포함한다. 본 방법은 합성 프레임들과 이미지들의 비교에 기초하여 입력 스케일 폐색 및 동작 인식 손실 함수를 사용하여 손실 값을 계산하는 단계를 포함한다. 본 방법은 손실 값에 기초하여 깊이 모델의 복수의 파라미터를 조정하는 단계를 포함한다. 훈련된 모델은 장면의 이미지를 수신하여서 상기 이미지에 따라 장면의 깊이 맵을 생성할 수 있다.

(52) CPC특허분류

G06T 7/97 (2017.01)

G06T 2207/10016 (2013.01)

G06T 2207/10028 (2013.01)

(72) 발명자

브로스토우, 가브리엘 제이.

미국, 캘리포니아주 94111, 샌프란시스코, 스위트
200, 원 페리 빌딩, 나이앤틱, 인크. 사내

피먼, 마이클

미국, 캘리포니아주 94111, 샌프란시스코, 스위트
200, 원 페리 빌딩, 나이앤틱, 인크. 사내

명세서

청구범위

청구항 1

프로세스에 의해 제작되는 훈련된 깊이 모델 및 명령어를 저장하는 비일시적 컴퓨터 판독 가능 저장 매체로서, 상기 프로세스는:

이미지들의 세트를 획득하는 단계로서, 상기 이미지들의 세트는 제1 타임 스탬프에서의 제1 이미지, 제2 타임스탬프에서의 제2 이미지 및 제3 타임스탬프에서의 제3 이미지를 포함하는 이미지들의 세트를 획득하는 단계;

상기 깊이 모델의 파라미터들에 기초하여 상기 제1 이미지 및 제3 이미지에 대한 깊이 맵들을 생성하기 위해 상기 깊이 모델에 적용하는 단계;

상기 제1 이미지 및 제3 이미지에 대한 포즈들을 생성하기 위하여 상기 제1 이미지 및 제3 이미지를 포즈 디코더에 적용하는 단계;

상기 제1 이미지 및 제3 이미지에 대한 포즈들과 깊이 맵들에 기초하여 합성 프레임들을 생성하는 단계로서, 제1 합성 프레임은 제1 이미지에 대한 포즈 및 깊이 맵을 기초로 상기 제2 타임스탬프에 대하여 생성되고, 제2 합성 프레임은 제3 이미지에 대한 포즈 및 깊이 맵을 기초로 상기 제2 타임스탬프에 대하여 생성되는 합성 프레임들을 생성하는 단계;

상기 합성 프레임들과 상기 제2 이미지의 비교에 기초하여 입력 스케일 폐색 인식 및 동작 인식 손실 함수를 사용하여 손실 값을 계산하는 단계로서,

제1 합성 프레임과 제2 이미지 사이의 제1 차이 및 제2 합성 프레임과 제2 이미지 사이의 제2 차이를 계산하는 단계; 및

상기 제1 차이 및 제2 차이 사이의 최소 차이를 식별하는 단계로서, 상기 손실 값은 상기 최소 차이에 기초하는, 상기 제1 차이 및 제2 차이 사이의 최소 차이를 식별하는 단계; 를 포함하는, 상기 손실 값을 계산하는 단계; 및

상기 손실 값에 기초하여 상기 깊이 모델의 파라미터들을 조정하는 단계를 포함하고, 상기 명령어는 컴퓨팅 장치에 의해 실행될 때 컴퓨팅 장치가 훈련된 깊이 모델을 장면의 이미지에 적용하여 장면의 깊이 맵을 생성하게 하는 비일시적 컴퓨터 판독 가능 저장 매체.

청구항 2

제1항에 있어서,

상기 이미지들의 세트는 단안 비디오를 포함하고, 상기 단안 비디오의 각각의 이미지는 대응하는 타임 스탬프에서 캡처되며, 상기 합성 프레임들을 생성하는 단계는,

상기 단안 비디오의 이미지들에 대해, 인접한 타임 스탬프들에서 합성 프레임들을 생성하는 단계를 포함하는 비일시적 컴퓨터 판독 가능 저장 매체.

청구항 3

제2항에 있어서,

상기 제1 차이 및 제2 차이를 계산하는 단계는,

상기 단안 비디오의 이미지들에 대해, 타임스탬프들이 일치하는 이미지들과 생성된 합성 프레임들 사이의 차이

들을 계산하는 단계를 포함하는 프로세스의 일부인 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 4

제1항에 있어서,

상기 제1 차이 및 제2 차이를 계산하는 단계는,

상기 제1 합성 프레임과 상기 제2 이미지 사이의 제1 세트의 차이 및 상기 제2 합성 프레임과 상기 제2 이미지 사이의 제2 세트의 차이를 계산하는 단계; 및

상기 최소 차이를 식별하는 단계로서, 상기 제2 이미지의 각 픽셀에 대해, 상기 제1 세트의 차이와 상기 제2 세트의 차이 사이의 최소 차이를 식별하는 단계를 포함하는 프로세스의 일부이며, 상기 손실 값은 상기 최소 차이에 기초하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 5

제3항에 있어서,

상기 입력 스케일 폐색 인식 및 동작 인식 손실 함수를 사용하여 손실 값을 계산하는 단계는,

상기 단안 비디오에서 하나 이상의 정적 특징을 식별하는 단계를 더 포함하고,

상기 손실 값은 상기 하나 이상의 정적 특징의 차이들을 필터링한 것에 기초하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 6

제5항에 있어서,

상기 단안 비디오에서 하나 이상의 정적 특징을 식별하는 단계는,

상기 단안 비디오의 제1 타임 스탬프에서의 제1 이미지의 영역 및 상기 단안 비디오의 제2 타임 스탬프에서의 제2 이미지의 영역을 식별하는 단계;

물체가 상기 제1 이미지와 상기 제2 이미지 사이의 유사한 위치에 있는지를 결정하는 단계; 및

물체가 상기 제1 이미지와 상기 제2 이미지 사이의 유사한 위치에 있다는 결정에 응답하여 상기 영역을 정적 특징으로 규정하는 단계를 포함하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 7

제1항에 있어서,

상기 이미지들의 세트는 제1 해상도에 있고, 상기 생성된 깊이 맵들은 상기 제1 해상도보다 낮은 제2 해상도에 있으며, 상기 깊이 모델을 제작하는 프로세스는,

상기 생성된 깊이 맵들을 상기 제2 해상도로부터 상기 제1 해상도로 업샘플링하는 단계를 더 포함하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 8

제1항에 있어서,

상기 이미지들의 세트는 입체 쌍의 카메라들에 의해 캡처된 각각의 입체 이미지 쌍을 갖는 입체 이미지 쌍들을

포함하고, 상기 합성 프레임들을 생성하는 단계는,

입체 이미지 쌍 각각에 대해, 입체 이미지 쌍의 제1 이미지에 기초하여 합성 프레임 생성하는 단계를 포함하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 9

제8항에 있어서,

상기 손실 값에 기초하여 상기 깊이 모델의 파라미터들을 조정하는 단계는,

입체 이미지 쌍 각각에 대해, 생성된 합성 프레임과 입체 이미지 쌍의 제2 이미지 사이의 차이를 계산하는 단계; 및

상기 파라미터들을 조정하여 차이들을 최소화하는 단계를 포함하는 비밀시적 컴퓨터 판독 가능 저장 매체.

청구항 10

컴퓨터 구현 방법으로서,

장면의 이미지를 수신하는 단계;

상기 이미지를 훈련된 깊이 모델에 입력하는 단계로서 상기 깊이 모델은,

이미지들의 세트를 획득하는 단계로서, 상기 이미지들의 세트는 제1 타임 스탬프에서의 제1 이미지, 제2 타임스탬프에서의 제2 이미지 및 제3 타임스탬프에서의 제3 이미지를 포함하는 이미지들의 세트를 획득하는 단계;

상기 깊이 모델의 파라미터들에 기초하여 제1 이미지 및 제3 이미지에 대한 깊이 맵들을 추출하기 위해 상기 제1 이미지 및 제3 이미지를 상기 깊이 모델에 입력하는 단계;

상기 제1 이미지 및 제3 이미지에 대한 포즈들을 추출하기 위해 상기 제1 이미지 및 제3 이미지를 포즈 디코더에 입력하는 단계;

상기 제1 이미지 및 제3 이미지에 대한 포즈들과 깊이 맵들에 기초하여 합성 프레임들을 생성하는 단계로서, 제1 합성 프레임은 제1 이미지에 대한 포즈 및 깊이 맵을 기초로 상기 제2 타임스탬프에 대하여 생성되고, 제2 합성 프레임은 3 이미지에 대한 포즈 및 깊이 맵을 기초로 상기 제2 타임스탬프에 대하여 생성되는 합성 프레임들을 생성하는 단계;

상기 합성 프레임들과 상기 제2 이미지의 비교에 기초하여 입력 스케일 폐색 인식 및 동작 인식 손실 함수를 사용하여 손실 값을 계산하는 단계로서, 상기 손실 값을 계산하는 단계는:

제1 합성 프레임과 제2 이미지 사이의 제1 차이 및 제2 합성 프레임과 제2 이미지 사이의 제2 차이를 계산하는 단계; 및

상기 제1 차이 및 제2 차이 사이의 최소 차이를 식별하는 단계로서, 상기 손실 값은 상기 최소 차이에 기초하는, 상기 제1 차이 및 제2 차이 사이의 최소 차이를 식별하는 단계; 를 포함하는, 상기 손실 값을 계산하는 단계; 및

상기 손실 값에 기초하여 상기 깊이 모델의 파라미터들을 조정하는 단계;를 포함하는 프로세스로 훈련되는 상기 이미지를 훈련된 깊이 모델에 입력하는 단계; 및

상기 훈련된 모델에 의해, 장면의 이미지에 대응하는 장면의 깊이 맵을 생성하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 11

제10항에 있어서,

상기 이미지들의 세트는 단안 비디오를 포함하고, 상기 단안 비디오의 각각의 이미지는 대응하는 타임 스탬프에서 캡처되며, 상기 합성 프레임들을 생성하는 단계는,

상기 단안 비디오의 이미지들에 대해, 인접한 타임 스탬프들에서 합성 프레임들을 생성하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 12

제11항에 있어서,

상기 제1 차이 및 제2 차이를 계산하는 단계는,

상기 단안 비디오의 이미지들에 대해, 타임스탬프들이 일치하는 이미지와 생성된 합성 프레임들 사이의 차이들을 계산하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 13

제10항에 있어서,

상기 제1 차이 및 제2 차이를 계산하는 단계는,

상기 제1 합성 프레임과 상기 제2 이미지 사이의 제1 세트의 차이 및 상기 제2 합성 프레임과 상기 제2 이미지 사이의 제2 세트의 차이를 계산하는 단계; 및

상기 최소 차이를 식별하는 단계로서, 상기 제2 이미지의 각 픽셀에 대해, 상기 제1 세트의 차이와 상기 제2 세트의 차이 사이의 최소 차이를 식별하는 단계; 를 포함하며, 상기 손실 값은 상기 최소 차이에 기초하는 컴퓨터 구현 방법.

청구항 14

제11항에 있어서,

상기 입력 스케일 폐색 인식 및 동작 인식 손실 함수를 사용하여 손실 값을 계산하는 단계는,

상기 단안 비디오에서 하나 이상의 정적 특징을 식별하는 단계를 더 포함하고, 상기 손실 값은 상기 하나 이상의 정적 특징의 차이들을 필터링한 것에 기초하는 컴퓨터 구현 방법.

청구항 15

제14항에 있어서,

상기 단안 비디오에서 하나 이상의 정적 특징을 식별하는 단계는,

상기 단안 비디오의 제1 타임 스탬프에서의 제1 이미지의 영역 및 상기 단안 비디오의 제2 타임 스탬프에서의 제2 이미지의 영역을 식별하는 단계;

물체가 상기 제1 이미지와 상기 제2 이미지 사이의 유사한 위치에 있는지를 결정하는 단계; 및

물체가 상기 제1 이미지와 상기 제2 이미지 사이의 유사한 위치에 있다는 결정에 응답하여 상기 영역을 정적 특징으로 규정하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 16

제10항에 있어서,

상기 이미지들의 세트는 제1 해상도에 있고, 상기 추출된 깊이 맵들은 상기 제1 해상도보다 낮은 제2 해상도에 있으며, 상기 깊이 모델을 훈련하는 프로세스는,

상기 추출된 깊이 맵들을 상기 제2 해상도로부터 상기 제1 해상도로 업샘플링하는 단계를 더 포함하는 컴퓨터 구현 방법.

청구항 17

제10항에 있어서,

상기 이미지들의 세트는 입체 쌍의 카메라들에 의해 캡처된 각각의 입체 이미지 쌍을 갖는 입체 이미지 쌍들을 포함하고, 상기 합성 프레임들을 생성하는 단계는,

입체 이미지 쌍 각각에 대해, 입체 이미지 쌍의 제1 이미지에 기초하여 합성 프레임을 생성하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 18

제17항에 있어서,

상기 손실 값에 기초하여 상기 모델의 파라미터들을 조정하는 단계는,

입체 이미지 쌍 각각에 대해, 생성된 합성 프레임과 입체 이미지 쌍의 제2 이미지 사이의 차이를 계산하는 단계; 및

상기 파라미터들을 조정하여 차이들을 최소화하는 단계를 포함하는 컴퓨터 구현 방법.

청구항 19

제10항에 있어서,

상기 장면의 깊이 맵에 기초하여 가상 콘텐츠로 증강된 장면의 이미지를 디스플레이하는 단계를 더 포함하는 컴퓨터 구현 방법.

청구항 20

비일시적 컴퓨터 판독 가능 저장 매체로서, 컴퓨팅 장치에 의해 실행될 때 컴퓨팅 장치로 하여금,

장면의 이미지를 수신하는 단계;

상기 이미지를 훈련된 깊이 모델에 입력하는 단계로서, 상기 깊이 모델은,

이미지들의 세트를 획득하는 단계;

깊이 인코더의 파라미터들에 기초하여 상기 이미지들에 대한 추상 깊이 특징들을 추출하기 위해 상기 이미지들을 상기 깊이 인코더에 입력하는 단계;

상기 이미지들에 대한 상기 추상 깊이 특징들을 연쇄 연결하는 단계;

상기 이미지들에 대한 포즈들을 추출하기 위해 상기 연쇄 연결된 추상 깊이 특징들을 포즈 디코더에 입력하는 단계;

상기 이미지들에 대한 포즈들과 상기 모델의 파라미터들에 기초하여 합성 프레임들을 생성하는 단계; 및

상기 합성 프레임들과 상기 이미지들의 세트의 비교에 기초하여 상기 모델의 파라미터들을 조정하는 단계를 포함하는 프로세스로 훈련되는 상기 이미지를 훈련된 깊이 모델에 입력하는 단계; 및

상기 훈련된 모델에 의해, 상기 장면의 이미지에 대응하는 장면의 깊이 맵을 생성하는 단계를 포함하는 작동들

을 수행하게 하는 명령어를 저장하는 비밀시적 컴퓨터 판독 가능 저장 매체.

발명의 설명

기술 분야

[0001] 설명된 기술적 요지는 일반적으로 단일 색상 입력 이미지로부터 깊이 맵을 추정하는 것에 관련된 것이고, 특히 비디오 데이터 및/또는 입체 이미지 데이터를 사용하여 훈련된 깊이 맵을 추정하기 위한 기계 학습 모델에 관련된 것이다.

배경 기술

[0002] 깊이 센싱은 내비게이션 및 장면 이해 모두에 적용된다. 많은 방법이 단일 색상 입력 이미지에서 깊이 맵을 결정하는 데에 훈련된 모델 또는 네트워크를 사용하고 있다. 깊이 추정 시스템을 훈련하기 위해 여러 가지 종류의 훈련 데이터를 사용하는 몇 가지 방법이 있다. 깊이 추정 시스템은 카메라에 의해 촬영된 동일한 장면의 이미지와 쌍을 이루는 환경(즉, 레이더(RADAR: radio detecting and ranging), 라이더(LIDAR: light detection and ranging) 등) 내의 물체에 대한 지상 실측 깊이(ground truth depth)를 확립하기 위해 감지 및 거리 측정 시스템을 사용하여 훈련시킬 수 있다. 감지 및 거리 측정 시스템은 물체의 지상 실측 깊이를 제공할 수 있지만, 많은 여러 가지 환경의 깊이를 감지하는 데 감지 및 거리 측정 시스템을 지속적으로 사용하는 것은 시간과 자원 측면에서 비용이 많이 드는 시도일 수 있다. 더욱이, 감지 및 거리 측정 시스템은 그 감지 및 거리 측정 시스템으로는 감지할 수 없게 하는 물질적 특성을 가질 수 있는 일부 물체(예를 들어, 반사성 물체)의 깊이는 결정할 수 없다.

[0003] 깊이 추정 시스템을 훈련하는 또 다른 방법은 동일한 장면의 입체 이미지 쌍들을 사용한다. 단일 시간 인스턴스에서 입체 이미지 쌍을 캡처하는 것은, 동일한 장면에 초점을 맞추고 있지만 어느 정도의 거리로 떨어져 위치한 두 대의 카메라를 사용하는 것에 의존한다. 깊이 추정 시스템은 입체 이미지 쌍의 입체 이미지들 중 한 입체 이미지에서 다른 입체 이미지로 투영하는 방식으로 작동한다. 깊이 추정 시스템은, 한 입체 이미지에서 다른 입체 이미지로의 투영을 위해, 입체 이미지 쌍을 촬영하는 데 사용되는 두 카메라의 물리적 위치들 간의 차이(disparity) - 척도화 깊이의 역(scaled inverse of depth) - 와 상대 변환 외에도 현재의 입체 이미지를 고려한다. 투영의 측광 재구성 오차를 캡처된 입체 이미지와 비교하여 최소화하는 데 있어서, 깊이 추정 시스템은 장면의 깊이를 결정할 수 있다.

[0004] 좀 더 새로운 방법은 끊임없이 변화하는 장면의 단안 비디오 데이터를 사용하여 깊이 추정 시스템을 훈련시킨다. 깊이 추정 시스템은 측광 재구성 오차를 최소화하면서 단안 비디오 데이터 중의 하나의 시간 이미지(temporal image)를 후속한 시간 이미지로 투영함으로써 훈련을 한다. 그러나 이러한 시스템은 하나의 시간 이미지로부터 다른 시간 이미지에 이르기까지의 뷰 안으로 들어오거나 그 뷰로부터 나가는 객체를 부정확하게 설명할 수 있는데, 이는 깊이 맵의 경계 주변과 깊이 맵 내에 아티팩트를 유발하게 된다. 또한, 현재까지는 기존 시스템은 깊이 업샘플링 아티팩트가 발생하기 쉬운 깊이 맵 업샘플링 전에 깊이 맵을 결정하기 위해 저해상도 이미지를 입력하고 있다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0005] 본 개시내용은 깊이 추정 모델을 훈련하고 사용하는 방법을 기술한다. 시스템은 깊이 추정 모델을 훈련시키기 위해 이미지들을 획득한다. 시스템은 그 이미지들을 깊이 모델에 입력하여서 깊이 모델의 파라미터들에 기초하여 각 이미지에 대한 깊이 맵을 추출한다. 시스템은 그 이미지들을 포즈 디코더에 입력하여 각 이미지에 대한 포즈를 추출한다. 시스템은 각 이미지에 대한 포즈와 깊이 맵에 기초하여 합성 프레임들을 생성한다. 시스템은 합성 프레임들과 이미지들의 비교에 기초하여 입력 스케일 폐쇄 및 동작 인식 손실 함수(input scale occlusion and motion aware loss function)를 사용하여 손실 값을 계산한다. 입력 스케일 폐쇄 및 동작 인식 손실 함수는 깊이 모델의 파라미터들을 정제하는 데 사용하기 위한 손실 값을 계산한다. 손실 함수에는 합성 프레임과 입력 이미지 사이의 픽셀마다의 측광 재구성 오차(photometric reconstruction error)의 계산이 포함된다. 손실 함수

는 또한 다른 이미지에 시간적으로 인접한 단안 비디오로부터의 시간 이미지들로부터 투영된 두 합성 프레임들 사이의 최소 측광 재구성 오차도 고려할 수 있다. 외관 정합 손실 계산에 영향을 주게 되는 합성 프레임 생성 중에, 업샘플링된 깊이 특징들(depth feature)도 사용될 수 있다. 손실 함수는 손실 값을 계산할 때 이미지들 중의 정적 특징들을 고려에 넣지 않은 생성된 마스크를 구현할 수도 있다. 시스템은 깊이 모델의 파라미터들을 손실 값에 기초하여 조정한다. 훈련된 모델에 의해, 장치는 장면의 이미지를 수신하여서 그 이미지에 따라 장면의 깊이 맵을 생성할 수 있다.

[0006] 일부 실시형태에서, 깊이 추정 모델은 단안 비디오를 포함하는 이미지 데이터로 훈련된다. 단안 비디오의 각 이미지는 각기 다른 시간에 캡처되어, 해당하는 타임 스탬프와 결합된다. 제1 타임 스탬프가 있는 제1 이미지와, 제2 타임 스탬프가 있는 제2 이미지를 사용한 예시적인 논의에서, 모델은 단안 비디오로부터의 제1 타임 스탬프가 있는 제1 이미지를 사용하여 제2 타임 스탬프에서 합성 프레임을 생성한다. 이 모델은 제2 타임 스탬프가 있는 제2 이미지와 합성 프레임 사이의 측광 재구성 오차를 계산한다. 이 모델은 인접한 타임 스탬프들을 갖는 단안 비디오로부터의 다른 이미지들의 쌍들을 사용하여 위의 절차를 따른다. 이 모델은 측광 재구성 오차를 최소화하기 위해 파라미터들을 조정한다. 다른 실시형태에서, 이 모델은 입체 이미지 쌍들을 포함하는 이미지 데이터로 훈련되며, 여기서 상기 입체 이미지 쌍 각각은 입체적 쌍의 카메라들에 의해 캡처된다. 이 모델은 합성 프레임을 생성할 때에, 입체 이미지 쌍으로부터 이미지들 중 하나(예를 들어, 왼쪽 이미지)를 취해서, 다른 이미지(예를 들어, 오른쪽 이미지)에서 합성 프레임을 생성한다. 이 모델은 상기 합성 프레임과 상기 다른 이미지 사이의 측광 재구성 오차를 계산한다. 이 모델은 측광 재구성 오차를 최소화하기 위해 파라미터들을 조정한다. 또 다른 실시형태에서, 이 모델은 단안 비디오 및 입체 이미지 쌍들 모두를 포함하는 이미지 데이터로 훈련된다.

도면의 간단한 설명

[0007] 도 1은 하나 이상의 실시형태에 따른 것으로, 네트워크 컴퓨팅 환경을 예시하고 있다.
 도 2는 하나 이상의 실시형태에 따른 것으로, 현실 세계와 대등한 지리를 갖는 가상 세계의 표현을 도시하고 있다.
 도 3은 하나 이상의 실시형태에 따른 것으로, 병렬 현실 게임의 예시적인 게임 인터페이스를 도시하고 있다.
 도 4는 하나 이상의 실시형태에 따른 것으로, 별도의 포즈 추정 모델을 사용하는 것과 깊이-포즈 하이브리드 모델을 사용하는 것을 대비한 개념적 비교를 예시하고 있다.
 도 5는 하나 이상의 실시형태에 따른 것으로, 깊이 추정 모델을 훈련하는 일반적인 프로세스를 설명하는 흐름도이다.
 도 6은 하나 이상의 실시형태에 따른 것으로, 깊이 추정 모델을 사용하는 일반적인 프로세스를 설명하는 흐름도이다.
 도 7은 하나 이상의 실시형태에 따른 것으로, 컴퓨팅 장치의 예시적인 아키텍처이다.
 도면 및 다음의 설명은 특정 실시형태들을 단지 예시로서 설명하는 것일 뿐이다. 본 기술분야의 숙련인은 구조 및 방법의 대안적인 실시형태들이 설명된 원리를 벗어나지 않고 사용될 수 있다는 것을 다음의 설명으로부터 쉽게 인식할 것이다. 이제 여러 실시형태들을 참조할 것이며, 그 예는 첨부 도면에 예시되어 있다.

발명을 실시하기 위한 구체적인 내용

[0008] 예시적인 위치 기반 병렬 현실 게임 시스템

[0009] 병렬 현실 게임은 현실 세계의 플레이어의 동작과 활동이 가상 세계에서의 활동에 영향을 미치도록 그리고 이와 반대의 경우도 이루어지도록 현실 세계 지리의 적어도 일부와 대등한 가상 세계 지리를 갖는 위치 기반 게임이다. 본 기술분야의 숙련인이 본원에 제공된 개시내용을 사용하여 이해하게 되는 바와 같이, 설명된 기술적 요지는 이미지 데이터로부터 깊이 정보를 결정하는 것이 바람직한 그 밖의 다른 상황에서 적용 가능하다. 또한, 컴퓨터 기반 시스템의 고유한 유연성은 시스템의 구성 요소들 간의 작업 및 기능을 아주 다양하게 구성, 조합, 및 분할할 수 있게 한다. 예를 들어, 본 개시내용의 양태들에 따른 시스템 및 방법은 단일 컴퓨팅 장치를 사용하여 구현되거나 또는 다수의 컴퓨팅 장치(예를 들어, 컴퓨터 네트워크에 연결된 것) 전반에 걸쳐 구현될 수 있다.

[0010] 도 1은 하나 이상의 실시형태에 따른 것으로, 네트워크 컴퓨팅 환경(100)을 예시하고 있다. 네트워크 컴퓨팅 환

경(100)은 현실 세계와 대등한 지리를 갖는 가상 세계에서의 플레이어의 상호 작용을 제공한다. 특히, 현실 세계의 지리적 영역이 가상 세계의 대응하는 영역에 직접 연결되거나 매핑될 수 있다. 플레이어는 현실 세계의 다양한 지리적 위치로 이동함으로써 가상 세계에서 이동할 수 있다. 예를 들어, 현실 세계에서의 플레이어의 위치를 추적하여서, 가상 세계에서의 플레이어의 위치를 업데이트하는 데 사용할 수 있다. 전형적으로, 현실 세계에서의 플레이어의 위치는 플레이어가 가상 세계와 상호 작용하는 통로인 클라이언트 장치(110)의 위치를 찾아서 플레이어가 동일한(또는 거의 동일한) 위치에 있다고 추정함으로써 결정된다. 예를 들어, 다양한 실시형태에서, 현실 세계에서의 플레이어의 위치가 가상 세계에서의 가상 요소의 가상 위치에 대응하는 현실 세계 위치의 임계 거리(예를 들어, 10 미터, 20 미터 등) 내에 있는 경우, 플레이어는 가상 요소와 상호 작용할 수 있다. 편의상, 다양한 실시형태들이 "플레이어의 위치"를 참조하여 설명되지만, 본 기술분야의 숙련인은 그러한 참조가 플레이어의 클라이언트 장치(110)의 위치를 가리킬 수 있다는 것을 이해할 것이다.

[0011] 이제, 일 실시형태에 따른 것으로, 별첨 현실 게임의 플레이어를 위한 게임 보드로서 작용할 수 있는 현실 세계(200)와 대등한 가상 세계(210)의 개념도를 도시하는 도 2를 참조한다. 예시된 바와 같이, 가상 세계(210)는 현실 세계(200)의 지리와 유사한 지형을 포함할 수 있다. 특히, 현실 세계(200)에서의 지리적 영역 또는 공간을 정의하는 좌표 범위는 가상 세계(210)에서의 가상 공간을 정의하는 대응하는 좌표 범위에 매핑된다. 현실 세계(200)에서의 좌표 범위는 타운, 이웃, 도시, 캠퍼스, 현장, 국가, 대륙, 지구 전체, 또는 기타 지리적 영역과 연관될 수 있다. 지리적 좌표 범위 내의 각 지리적 좌표는 가상 세계의 가상 공간 내의 대응하는 좌표에 매핑된다.

[0012] 가상 세계(210)에서의 플레이어의 위치는 현실 세계(200)에서의 플레이어의 위치에 대응한다. 예를 들어, 현실 세계(200)에서 위치(212)에 위치한 플레이어 A는 가상 세계(210)에서의 대응하는 위치(222)를 갖는다. 유사하게, 현실 세계에서 위치(214)에 위치한 플레이어 B는 가상 세계에서의 대응하는 위치(224)를 갖는다. 플레이어가 현실 세계에서 지리적 좌표 범위 내에서 이동함에 따라, 플레이어는 가상 세계(210)에서 가상 공간을 정의하는 좌표 범위 내에서도 이동한다. 특히, 플레이어가 휴대하는 모바일 컴퓨팅 장치와 연관된 위치 확인 시스템(예를 들어, GPS 시스템)은 플레이어가 현실 세계에서 지리적 좌표 범위를 탐색할 때 플레이어의 위치를 추적하는 데 사용될 수 있다. 현실 세계(200)에서의 플레이어의 위치와 연관된 데이터는 가상 세계(210)에서의 가상 공간을 정의하는 좌표의 대응하는 범위 내의 플레이어의 위치를 업데이트하는 데 사용된다. 이러한 방식에서, 플레이어는 현실 세계의 특정 개별 위치(200)에서 위치 정보를 확인하거나 주기적으로 업데이트하지 않고 현실 세계(200)에서 대응하는 지리적 좌표 범위 사이에서 단지 이동하기만 해도 가상 세계(210)에서의 가상 공간을 정의하는 좌표 범위 내에서 연속 트랙을 따라 탐색할 수 있다.

[0013] 위치 기반 게임은 플레이어가 가상 세계의 다양한 가상 위치에 흩어져 있는 다양한 가상 요소들 및/또는 가상 물체들로 이동하고/하거나 그와 상호 작용하는 것을 요구하는 복수의 게임 목표를 포함할 수 있다. 플레이어는 현실 세계의 가상 요소들 또는 물체들의 대응하는 위치로 이동함으로써 그러한 가상 위치로 이동할 수 있다. 예를 들어, 위치 확인 시스템은, 플레이어가 현실 세계를 계속해서 탐색함에 따라 대등한 가상 세계도 계속해서 탐색할 수 있도록, 플레이어의 위치를 계속해서 추적할 수 있다. 그런 다음 플레이어는 특정 위치에서 다양한 가상 요소들 및/또는 물체들과 상호 작용하여 하나 이상의 게임 목표를 달성하거나 수행할 수 있다.

[0014] 예를 들어, 한 가지 게임 목표는 플레이어들이 가상 세계(210)의 다양한 가상 위치들에 위치한 가상 요소들(230)과 상호 작용하게 하는 것이다. 이러한 가상 요소들(230)은 현실 세계(200)의 랜드마크들, 지리적 위치들, 또는 물체들(240)에 연결될 수 있다. 현실 세계의 랜드마크 또는 물체(240)는 예술 작품, 기념물, 건물, 상점, 도서관, 박물관, 또는 기타 적절한 현실 세계의 랜드마크 또는 물체일 수 있다. 상호 작용에는 포획, 소유권 주장, 일부 가상 아이템 사용, 일부 가상 화폐 사용 등이 포함된다. 플레이어는 이러한 가상 요소들(230)을 포획하기 위해서는 현실 세계에서 가상 요소들(230)에 연결된 랜드마크 또는 지리적 위치(240)로 이동해야 하고, 가상 세계(210)에서 가상 요소들(230)과 임의의 필요한 상호 작용을 수행해야 한다. 예를 들어, 도 2의 플레이어 A는 특정 랜드마크(240)와 연결된 가상 요소(230)와 상호 작용하거나 또는 그것을 포획하기 위해 현실 세계(200)에서 랜드마크(240)로 이동해야 할 수 있다. 가상 요소(230)와의 상호 작용에는 사진 촬영 및/또는 가상 요소(230)와 관련된 랜드마크 또는 물체(240)에 대한 기타 정보를 확인, 획득, 또는 포획하는 것과 같은 현실 세계에서의 활동이 필요할 수 있다.

[0015] 게임 목표는 플레이어가 위치 기반 게임에서 그 플레이어가 수집한 하나 이상의 가상 아이템을 사용할 것을 요구할 수 있다. 예를 들어, 플레이어는 게임 목표를 완료하는 데 유용할 수 있는 가상 아이템(예를 들어, 무기, 생물, 파워 증강, 또는 기타 아이템)을 찾으려고 가상 세계(210)를 이동할 수 있다. 이러한 가상 아이템은 현실 세계(200)에서 다른 위치로 이동하거나 가상 세계(210) 또는 현실 세계(200)에서 다양한 활동을 완료함으로써

찾거나 수집할 수 있다. 도 2에 도시된 예에서, 플레이어는 하나 이상의 가상 요소(230)를 포획하기 위해 가상 아이템(232)을 사용한다. 특히, 플레이어는 가상 세계(210) 내의 위치들에 있는 가상 아이템들(232)을 가상 요소(230) 내이나 또는 그에 근접하게 배치할 수 있다. 하나 이상의 가상 아이템(232)을 이러한 방식으로 배치하는 것은 특정 플레이어 또는 특정 플레이어의 팀/진영이 가상 요소(230)를 포획한 결과가 될 수 있다.

[0016] 특정의 일 구현 예에서, 플레이어는 병렬 현실 게임의 일부로서 가상 에너지를 수집해야 할 수 있다. 도 2에 도시된 바와 같이, 가상 에너지(250)는 가상 세계(210)의 여러 위치에 흩어져 있을 수 있다. 플레이어는 실제 세계(200)에서 가상 에너지(250)의 대응 위치로 이동하여 가상 에너지(250)를 수집할 수 있다. 가상 에너지(250)는 게임에서 가상 아이템에 파워를 공급하고/하거나 다양한 게임 목표를 수행하는 데 사용될 수 있다. 모든 가상 에너지(250)를 잃은 플레이어는 게임에서 끊길 수 있다.

[0017] 본 개시내용의 양태들에 따르면, 병렬 현실 게임은 게임의 모든 참가자가 동일한 가상 세계를 공유하는 대규모 멀티플레이어 위치 기반 게임일 수 있다. 플레이어들은 별도의 팀 또는 진영으로 나누어질 수 있으며, 가상 요소를 포획하거나 가상 요소의 소유권을 주장하는 것과 같은 한 가지 이상의 게임 목표를 달성하기 위해 함께 노력할 수 있다. 이러한 방식으로, 병렬 현실 게임은, 본질적으로, 게임 내에서 플레이어들 간의 협력을 장려하는 소셜 게임이 될 수 있다. 상대 팀의 플레이어들은 병렬 현실 게임 중에 서로 대결(또는 때로는 상호 목표를 달성하기 위해 협력)할 수 있다. 한 플레이어가 가상 아이템을 사용하여 상대 팀의 플레이어들을 공격하거나 그 플레이어들의 진행을 방해할 수 있다. 일부 경우에는, 플레이어들을 병렬 현실 게임에서의 협력 또는 상호 작용 이벤트를 위해 현실 세계의 위치에 모이도록 권장한다. 이러한 경우, 게임 서버는 플레이어들이 사실상 실제로 존재하고 속임수가 아니라는 것을 보장하려는 노력을 한다.

[0018] 병렬 현실 게임은 그 병렬 현실 게임 내에서 게임 플레이를 향상시키고 장려하기 위한 다양한 특징들(feature)을 가질 수 있다. 예를 들어, 플레이어는 게임 전반에 걸쳐 (예를 들어, 게임 내 아이템 구매, 다른 아이템 회수, 아이템 만들기 등을 위해) 사용할 수 있는 가상 화폐 또는 다른 가상 보상(예를 들어, 가상 토큰, 가상 포인트, 가상 물질 자원 등)을 축적할 수 있다. 플레이어는 한 가지 이상의 게임 목표를 완료하여 게임 내에서 경험을 얻음에 따라 다양한 레벨로 발전해나갈 수 있다. 일부 실시형태에서, 플레이어들은 게임에서 제공되는 하나 이상의 통신 인터페이스를 통해 서로 통신할 수 있다. 플레이어들은 또한 게임 내에서 게임 목표를 완료하는 데 사용할 수 있는 증강된 "파워" 또는 가상 아이템을 얻을 수도 있다. 본 기술분야의 숙련인이 본원에 제공된 개시내용을 사용하여 당연히 이해하게 되는 바와 같이, 그 밖의 다른 다양한 다른 게임 특징들(game feature)이 본 개시내용의 범위를 벗어나지 않고 병렬 현실 게임에 포함될 수 있다.

[0019] 다시 도 1을 참조하면, 네트워크 컴퓨팅 환경(100)은 게임 서버(120)가 네트워크(105)를 통해 클라이언트 장치(110)와 통신하여 클라이언트 장치(110)에서 플레이어들에게 병렬 현실 게임을 제공하는 클라이언트-서버 아키텍처를 사용한다. 네트워크 컴퓨팅 환경(100)은 또한 스폰서/광고주 시스템 또는 비즈니스 시스템과 같은 기타 외부 시스템을 포함할 수 있다. 도 1에는 단지 하나의 클라이언트 장치(110)만 도시되어 있지만, 임의의 수의 클라이언트(110) 또는 기타 외부 시스템이 네트워크(105)를 통하여 게임 서버(120)에 연결될 수 있다. 더욱이, 네트워크화된 컴퓨팅 환경(100)은 상이하거나 추가적인 요소들을 포함할 수 있고, 기능은 후술하는 것과는 다른 방식으로 클라이언트 장치(110)와 서버(120) 사이에 분산될 수 있다.

[0020] 클라이언트 장치(110)는 플레이어가 게임 서버(120)와 인터페이스하기 위해 사용할 수 있는 임의의 휴대용 컴퓨팅 장치일 수 있다. 예를 들어, 클라이언트 장치(110)는 무선 장치, 개인 정보 단말기(PDA), 휴대용 게임 장치, 휴대 전화기, 스마트폰, 태블릿, 내비게이션 시스템, 휴대용 GPS 시스템, 웨어러블 컴퓨팅 장치, 하나 이상의 프로세서를 갖춘 디스플레이, 또는 기타 이러한 장치일 수 있다. 다른 예로서, 클라이언트 장치(110)는 데스크톱 또는 랩톱 컴퓨터와 같은 종래의 컴퓨터 시스템을 포함한다. 또한, 클라이언트 장치(110)는 컴퓨팅 장치를 갖춘 차량일 수 있다. 요컨대, 클라이언트 장치(110)는 플레이어가 게임 서버(120)와 상호 작용할 수 있게 하는 임의의 컴퓨터 장치 또는 시스템일 수 있다. 컴퓨팅 장치로서의 클라이언트 장치(110)는 하나 이상의 프로세서 및 하나 이상의 컴퓨터 판독 가능 저장 매체를 포함할 수 있다. 컴퓨터 판독 가능 저장 매체는 프로세서에게 작업을 수행하게 하는 명령을 저장할 수 있다. 클라이언트 장치(110)는 스마트폰 또는 태블릿과 같이 플레이어가 쉽게 휴대하거나 또는 탄 방법으로 운반할 수 있는 휴대용 컴퓨팅 장치인 것이 바람직하다.

[0021] 클라이언트 장치(110)는 게임 서버(120)에 물리적 환경의 감각 데이터를 제공하는 게임 서버(120)와 통신한다. 클라이언트 장치(110)는 그 클라이언트 장치(110)가 있는 물리적 환경 내의 장면의 2차원 이미지 데이터를 캡처하는 카메라 어셈블리(125)를 포함한다. 클라이언트 장치(110)는 또한 예를 들어 게임 서버(120)에 의해 훈련되는 기계 학습 모델인 깊이 추정 모델(130)을 포함한다. 도 1에 도시된 실시형태에서, 각각의 클라이언트 장치

(110)는 게임 모듈(135) 및 위치 확인 모듈(140)과 같은 소프트웨어 구성 요소를 포함한다. 클라이언트 장치(110)는 플레이어로부터 정보를 수신하고/하거나 플레이어에게 정보를 제공하기 위한 다양한 기타 입출력 장치를 포함할 수 있다. 입력/출력 장치의 예에는 디스플레이 스크린, 터치스크린, 터치패드, 데이터 입력 키, 스피커, 및 음성 인식에 적합한 마이크가 포함된다. 클라이언트 장치(110)는 또한 동작 센서, 가속도계, 자이로스코프, 기타 관성 측정 유닛(IMU), 기압계, 위치 확인 시스템, 온도계, 광 센서 등을 포함하지만 이에 제한되지 않는 클라이언트 장치(110)로부터의 데이터를 기록하기 위한 기타 다양한 센서를 포함할 수 있다. 클라이언트 장치(110)는 네트워크(105)를 통하여 통신을 제공하기 위한 네트워크 인터페이스를 추가로 포함할 수 있다. 네트워크 인터페이스는, 예를 들어 송신기, 수신기, 포트, 컨트롤러, 안테나, 또는 기타 적절한 구성 요소를 포함하여, 하나 이상의 네트워크와 인터페이스하기 위한 임의의 적절한 구성 요소를 포함할 수 있다.

[0022] 카메라 어셈블리(125)는 클라이언트 장치(110)가 있는 환경의 장면의 이미지 데이터를 캡처한다. 카메라 어셈블리(125)는 다양한 캡처 레이트에서 다양한 컬러 캡처 범위를 갖는 다양하게 다른 포토 센서들을 사용할 수 있다. 카메라 어셈블리(125)는 광각 렌즈 또는 망원 렌즈를 포함할 수 있다. 카메라 어셈블리(125)는 단일 이미지 또는 비디오를 이미지 데이터로서 캡처하도록 구성될 수 있다. 추가적으로, 카메라 어셈블리(125)의 배향은 그 카메라 어셈블리(125)가 수평선을 향하도록 하여 지면과 평행할 수 있다. 카메라 어셈블리(125)는 이미지 데이터를 캡처하고, 그 이미지 데이터를 클라이언트 장치(110) 상의 컴퓨팅 장치와 공유한다. 이미지 데이터에는 감각 데이터(예를 들어, 환경의 온도, 밝기) 또는 캡처 데이터(예를 들어, 노출, 파도함, 셔터 속도, 초점 거리, 캡처 시간 등)를 포함한 이미지 데이터의 기타 세부 정보를 설명하는 메타 데이터가 추가될 수 있다. 카메라 어셈블리(125)는 이미지 데이터를 캡처할 수 있는 하나 이상의 카메라를 포함할 수 있다. 일 예에서, 카메라 어셈블리(125)는 하나의 카메라를 포함하며, 단안 이미지 데이터를 캡처하도록 구성된다. 다른 예에서, 카메라 어셈블리(125)는 2개의 카메라를 포함하며, 입체 이미지 데이터를 캡처하도록 구성된다. 다양한 그 밖의 다른 구현 예에서, 카메라 어셈블리(125)는, 각각 이미지 데이터를 캡처하도록 구성된 복수의 카메라를 포함한다.

[0023] 깊이 추정 모델(130)은 장면의 입력 이미지를 수신하고, 그 입력 이미지에 기초하여 장면의 깊이를 출력한다. 깊이 추정 모델(130)은 깊이 추정 훈련 시스템에 의해 훈련되며, 아래에서 더 상세히 논의되는 깊이 추정 훈련 시스템에 의해 업데이트되거나 조정될 수 있다. 수신된 입력 이미지는 카메라 어셈블리(125)의 카메라에 의해서, 또는 다른 클라이언트 장치(110)로부터의 다른 카메라에 의해 캡처될 수 있다. 일부 실시형태에서, 수신된 입력 이미지는 입력 이미지의 본질을 특징하는, 이미지에 첨부된 메타 데이터를 갖는다. 이미지의 본질은 이미지를 캡처할 때의 카메라의 하나 이상의 기하학적 속성, 예를 들어, 이미지를 캡처할 때의 카메라의 초점 거리, 카메라의 주점 오프셋(principal point offset), 카메라의 비대칭도(skew) 등을 나타낸다. 이러한 본질을 사용하여, 깊이 추정 모델(130)은 그 본질을 설명하는 내부 행렬(intrinsic matrix)을 생성할 수 있다. 일부 실시형태에서, 깊이 추정 모델(130)은 입력 이미지가 임계 해상도를 초과하는지를 결정한다. 그렇지 않은 경우, 깊이 추정 모델(130)은 장면의 깊이 맵을 결정하기 전에 입력 이미지를 원하는 해상도로 업샘플링할 수 있다. 깊이 추정 모델(130)은 이미지를 (수신된 대로 또는 업샘플링한 후에) 입력하여, 장면의 깊이 맵을 결정한다. 기계 학습 알고리즘은 훈련 및/또는 추론을 위해 깊이 추정 모델(130)에서 구현될 수 있다.

[0024] 게임 모듈(135)은 플레이어에게 병렬 현실 게임에 참여할 수 있는 인터페이스를 제공한다. 게임 서버(120)는 그 게임 서버(120)로부터 멀리 떨어진 위치에 있는 플레이어들에게 게임의 로컬 버전을 제공하기 위해 클라이언트 장치(110)에 있는 게임 모듈(135)이 사용하도록 하기 위한 게임 데이터를 네트워크(105)를 통하여 클라이언트 장치(110)로 전송한다. 게임 서버(120)는 네트워크(105)를 통한 통신을 제공하기 위한 네트워크 인터페이스를 포함할 수 있다. 네트워크 인터페이스는, 예를 들어 송신기, 수신기, 포트, 컨트롤러, 안테나, 또는 기타 적절한 구성 요소를 포함하여, 하나 이상의 네트워크와 인터페이스하기 위한 임의의 적절한 구성 요소를 포함할 수 있다.

[0025] 클라이언트 장치(110)에 의해 실행되는 게임 모듈(135)은 플레이어와 병렬 현실 게임 간의 인터페이스를 제공한다. 게임 모듈(135)은, 게임과 연관된 가상 세계를 디스플레이하며(예를 들어, 가상 세계의 이미지를 렌더링하고) 사용자가 다양한 게임 목표를 수행하도록 가상 세계에서 상호 작용할 수 있게 하는, 클라이언트 장치(110)와 연관된 디스플레이 장치에 사용자 인터페이스를 제공할 수 있다. 일부 다른 실시형태에서, 게임 모듈(135)은 병렬 현실 게임으로부터의 가상 요소들로 증강된 현실 세계로부터의 (예를 들어, 카메라 어셈블리(125)에 의해 캡처된) 이미지 데이터를 제공한다. 이러한 실시형태에서, 게임 모듈(135)은 클라이언트 장치의 다른 구성 요소들로부터 수신된 다른 정보에 따라 가상 콘텐츠를 생성하고/하거나 가상 콘텐츠를 조정할 수 있다. 예를 들어, 게임 모듈(135)은 이미지 데이터에서 캡처된 장면의 깊이 맵(예를 들어, 깊이 추정 모델(130)에 의해 결정됨)에 따라 사용자 인터페이스에 디스플레이될 가상 물체를 조정할 수 있다.

- [0026] 게임 모듈(135)은 또한 플레이어가 디스플레이 화면을 보지 않아도 게임과 상호 작용할 수 있게 하는 다양한 다른 출력을 제어할 수 있다. 예를 들어, 게임 모듈(135)은 플레이어가 디스플레이 화면을 보지 않고도 게임을 할 수 있게 하는 다양한 오디오, 진동, 또는 기타 알림을 제어할 수 있다. 게임 모듈(135)은 게임의 정확한 표현을 사용자에게 제공하기 위해, 게임 서버(120)로부터 수신된 게임 데이터에 액세스할 수 있다. 게임 모듈(135)은 플레이어의 입력을 수신 및 처리하여, 업데이트를 네트워크(105)를 통해 게임 서버(120)에 제공할 수 있다. 게임 모듈(135)은 또한 클라이언트 장치(110)에 의해 디스플레이될 게임 콘텐츠를 생성 및/또는 조정할 수 있다. 예를 들어, 게임 모듈(135)은 깊이 정보(예를 들어, 깊이 추정 모델(130)에 의해 결정됨)에 기초하여 가상 요소를 생성할 수 있다.
- [0027] 위치 확인 모듈(140)은 클라이언트 장치(110)의 위치를 모니터링하기 위한 임의의 장치 또는 회로일 수 있다. 예를 들어, 위치 확인 모듈(140)은 위성 항법 위치 확인 시스템(예를 들어, GPS 시스템, 갈릴레오(Galileo) 위치 확인 시스템, 범 지구 위성 항법 시스템(GLONASS), 베이더우(BeiDou) 위성 항법 및 위치 확인 시스템), 관성 항법 시스템, 셀룰러 타워 또는 Wi-Fi 핫스팟에 대한 삼각 측량 및/또는 근접도를 사용하여 IP 주소에 기초하는 추측 항법 시스템, 및/또는 위치를 결정하기 위한 기타 적합한 기술을 사용하여 실제 또는 상대적 위치를 결정할 수 있다. 위치 결정 모듈(140)은 클라이언트 장치(110) 위치를 정확하게 위치 결정하는 데 도움이 될 수 있는 그 밖의 다른 다양한 센서를 추가로 포함할 수 있다.
- [0028] 플레이어가 현실 세계에서 클라이언트 장치(110)와 함께 돌아다님에 따라, 위치 결정 모듈(140)은 플레이어의 위치를 추적하여 플레이어 위치 정보를 게임 모듈(135)에 제공한다. 게임 모듈(135)은 현실 세계에서의 플레이어의 실제 위치에 기초하여 게임과 연관된 가상 세계에서의 플레이어 위치를 업데이트한다. 따라서, 플레이어는 현실 세계에서 클라이언트 장치(110)를 단지 휴대하거나 운반하기만 해도 가상 세계와 상호 작용할 수 있다. 특히, 가상 세계에서의 플레이어의 위치는 현실 세계에서의 플레이어의 위치에 대응할 수 있다. 게임 모듈(135)은 플레이어 위치 정보를 네트워크(105)를 통해서 게임 서버(120)에 제공할 수 있다. 게임 서버(120)는 그에 응답하여, 부정 행위자가 클라이언트 장치(110)의 위치를 속이는 것을 방지하기 위해 클라이언트 장치(110)의 위치를 확인하는 다양한 기술을 규정할 수 있다. 플레이어와 연관된 위치 정보는, 플레이어의 위치 정보에 접근이 있을 것이라는 점과 그 위치 정보가 게임과 관련해서 (예를 들어, 가상 세계에서의 플레이어 위치를 업데이트하기 위해) 어떻게 활용되는지를 플레이어에게 통지한 후에 허락이 승인된 경우에만, 활용될 수 있다는 것을 이해해야 한다. 또한, 플레이어와 연관된 모든 위치 정보는 플레이어의 개인 정보를 보호하는 방식으로 저장되고 유지될 것이다.
- [0029] 게임 서버(120)는 임의의 컴퓨팅 장치일 수 있고, 하나 이상의 프로세서 및 하나 이상의 컴퓨터 판독 가능 저장 매체를 포함할 수 있다. 컴퓨터 판독 가능 저장 매체는 프로세서에게 작업을 수행하게 하는 명령을 저장할 수 있다. 게임 서버(120)는 게임 데이터베이스(115)를 포함하거나 또는 그와 통신할 수 있다. 게임 데이터베이스(115)는 네트워크(105)를 통해 클라이언트(들)(120)에게 공여 또는 제공될 병렬 현실 게임에서 사용되는 게임 데이터를 저장한다.
- [0030] 게임 데이터베이스(115)에 저장된 게임 데이터는, (1) 병렬 현실 게임에서의 가상 세계와 연관된 데이터(예를 들어, 디스플레이 장치에서 가상 세계를 렌더링하는 데 사용되는 이미지 데이터, 가상 세계에서의 위치들의 지리적 좌표 등); (2) 병렬 현실 게임의 플레이어와 연관된 데이터(예를 들어, 플레이어 정보, 플레이어 경험 수준, 플레이어 통화, 가상 세계/현실 세계에서의 현재 플레이어 위치, 플레이어 에너지 수준, 플레이어 선호도, 팀 정보, 진영 정보 등을 포함하지만 이에 한정되지 않는 플레이어 프로필); (3) 게임 목표와 연관된 데이터(예를 들어, 현재 게임 목표, 게임 목표의 상태, 과거 게임 목표, 미래 게임 목표, 원하는 게임 목표 등과 연관된 데이터); (4) 가상 세계의 가상 요소들과 연관된 데이터(예를 들어, 가상 요소들의 위치, 가상 요소들의 유형, 가상 요소들과 연관된 게임 목표, 가상 요소들에 대한 대응하는 현실 세계 위치 정보, 가상 요소들의 습성, 가상 요소들의 관련성 등); (5) 가상 세계 요소들과 결부된 현실 세계 물체, 랜드마크, 위치와 연관된 데이터(예를 들어, 현실 세계 물체/랜드마크의 위치, 현실 세계 물체/랜드마크에 대한 설명, 현실 세계 물체에 결부된 가상 요소들의 관련성 등); (6) 게임 상태(예를 들어, 현재 플레이어 수, 게임 목표의 현재 상태, 플레이어 순위표 등); (7) 플레이어의 동작/입력과 연관된 데이터(예를 들어, 현재 플레이어 위치, 과거 플레이어 위치, 플레이어의 이동, 플레이어의 입력, 플레이어의 질문, 플레이어의 대화 등); 및 (8) 병렬 현실 게임을 구현하는 중에 사용되거나, 관련되거나, 또는 획득된 기타 데이터를 포함할 수 있다. 게임 데이터베이스(115)에 저장된 게임 데이터는 시스템 관리자에 의해, 그리고/또는 시스템(100)의 사용자/플레이어로부터 수신된, 예컨대 네트워크(105)를 통해서 클라이언트 장치(110)로부터 수신된 데이터에 의해, 오프라인이나 또는 실시간으로 채워질 수 있다.

- [0031] 게임 서버(120)는 클라이언트 장치(110)로부터(예를 들어, 원격 프로시저 호출(RPC)을 통해) 게임 데이터 요청을 수신하고 네트워크(105)를 통해 그 요청에 응답하도록 구성될 수 있다. 예를 들어, 게임 서버(120)는 게임 데이터를 하나 이상의 데이터 파일로 인코딩하여, 그 데이터 파일을 클라이언트 장치(110)에 제공할 수 있다. 또한, 게임 서버(120)는 게임 데이터(예를 들어, 플레이어 위치, 플레이어 동작, 플레이어 입력 등)를 클라이언트 장치(110)로부터 네트워크(105)를 통해 수신하도록 구성될 수 있다. 예를 들어, 클라이언트 장치(110)는, 게임 서버(120)가 게임을 위한 임의의 모든 변경된 조건을 반영하기 위해 게임 데이터베이스(115)의 게임 데이터를 업데이트하는 데 사용하는, 플레이어 입력 및 기타 업데이트를 게임 서버(120)에 주기적으로 전송하도록 구성될 수 있다.
- [0032] 도시된 실시형태에서, 서버(120)는 범용 게임 모듈(145), 상용 게임 모듈(150), 데이터 수집 모듈(155), 이벤트 모듈(160), 및 깊이 추정 훈련 시스템(170)을 포함한다. 위에서 언급한 바와 같이, 게임 서버(120)는 게임 서버(120)의 일부이거나 또는 원격으로 액세스될 수 있는 게임 데이터베이스(115)와 상호 작용한다(예를 들어, 게임 데이터베이스(115)는 네트워크(105)를 통해 액세스되는 분산 데이터베이스일 수 있다). 다른 실시형태에서, 게임 서버(120)는 상이한 및/또는 추가 요소들을 포함한다. 또한, 기능들이 설명된 것과 다른 방식으로 요소들 사이에 분산될 수 있다. 예를 들어, 게임 데이터베이스(115)는 게임 서버(120)에 통합될 수 있다.
- [0033] 범용 게임 모듈(145)은 모든 플레이어를 위한 병렬 현실 게임을 호스팅하고, 모든 플레이어를 위한 병렬 현실 게임의 현재 상태에 대한 권한을 갖는 소스 역할을 한다. 호스트로서의 범용 게임 모듈(145)은 플레이어에게 표시하기 위한 게임 콘텐츠를 예를 들어 각각의 클라이언트 장치(110)를 통해 생성한다. 범용 게임 모듈(145)은 병렬 현실 게임을 호스팅할 때 게임 데이터를 검색 및/또는 저장하기 위해 게임 데이터베이스(115)에 액세스할 수 있다. 범용 게임 모듈(145)은 또한 클라이언트 장치(110)로부터 게임 데이터(예를 들어, 깊이 정보, 플레이어 입력, 플레이어 위치, 플레이어 동작, 랜드마크 정보 등)를 수신하고, 수신된 게임 데이터를 병렬 현실 게임의 모든 플레이어를 위한 전체 병렬 현실 게임에 통합시킨다. 범용 게임 모듈(145)은 또한 네트워크(105)를 통한 클라이언트 장치(110)로의 게임 데이터 전달도 관리할 수 있다. 범용 게임 모듈(145)은 또한, 클라이언트 장치(110)와 게임 서버(120) 간의 연결 보증, 다양한 클라이언트 장치(110) 간의 연결 확립, 및 다양한 클라이언트 장치(110)의 위치 확인을 포함하지만 이에 제한되지 않는 클라이언트 장치(110)의 보안 측면도 관리할 수 있다.
- [0034] 상용 게임 모듈 하나가 포함되는 실시형태에서, 상용 게임 모듈(150)은 범용 게임 모듈(145)과 별개이거나 범용 게임 모듈의 일부일 수 있다. 상용 게임 모듈(150)은 현실 세계에서의 상업 활동과 결부된 다양한 게임 특징들을 병렬 현실 게임 내에 포함시키는 것을 관리할 수 있다. 예를 들어, 상용 게임 모듈(150)은 스폰서/광고주, 비즈니스, 또는 기타 엔티티와 같은 외부 시스템으로부터의 요청을 네트워크(105)를 통해(네트워크 인터페이스를 통해) 수신하여서 상업 활동과 결부된 게임 특징들을 병렬 현실 게임 내에 포함시킬 수 있다. 그런 다음, 상용 게임 모듈(150)은 이러한 게임 특징들을 병렬 현실 게임에 포함시키도록 할 수 있다.
- [0035] 게임 서버(120)는 데이터 수집 모듈(155)을 추가로 포함할 수 있다. 데이터 수집 모듈 하나가 포함되는 실시형태에서, 데이터 수집 모듈(155)은 범용 게임 모듈(145)과 별개이거나 범용 게임 모듈의 일부일 수 있다. 데이터 수집 모듈(155)은 현실 세계에서의 데이터 수집 활동과 결부된 다양한 게임 특징들을 병렬 현실 게임 내에 포함시키는 것을 관리할 수 있다. 예를 들어, 데이터 수집 모듈(155)은 데이터 수집 활동과 결부된 게임 특징들을 병렬 현실 게임에 포함시키기 위해, 게임 데이터베이스(115)에 저장된 게임 데이터를 수정할 수 있다. 데이터 수집 모듈(155)은 또한 데이터 수집 활동에 따라 플레이어가 수집한 데이터를 분석하고 그 데이터를 다양한 플랫폼에서 접근할 수 있도록 제공할 수 있다.
- [0036] 이벤트 모듈(160)은 병렬 현실 게임에서 플레이어가 이벤트에 액세스하는 것을 관리한다. "이벤트"라는 용어는 편의상 사용되지만, 특정 위치 또는 시간의 특정 이벤트를 지칭할 필요는 없음을 이해해야 한다. 오히려 이 용어는 플레이어가 해당 콘텐츠에 액세스할 수 있는지 여부를 결정하는 데 하나 이상의 액세스 기준이 사용되는 액세스 제어식 게임 콘텐츠를 제공하는 것은 그 어떤 것도 지칭할 수 있다. 이러한 콘텐츠는 액세스 제어가 적거나 없는 게임 콘텐츠를 포함하는 대형 병렬 현실 게임의 일부일 수 있거나, 독립형 액세스 제어식 병렬 현실 게임일 수 있다.
- [0037] 깊이 추정 훈련 시스템(170)은 깊이 추정 모델, 예를 들어 클라이언트 장치(110)에 제공된 깊이 추정 모델(130)을 훈련시킨다. 깊이 추정 훈련 시스템(170)은 깊이 추정 모델을 훈련시키는 데 사용하기 위한 이미지 데이터를 수신한다. 일반적으로 말하면, 깊이 추정 훈련 시스템(170)은 이미지 데이터를 처리하고, 이미지 데이터를 깊이 추정 모델과 포즈 추정 모델에 입력하고, 이미지를 다른 이미지 상에 합성 프레임으로서 투영하고, 깊이

추정 모델의 파라미터들을 반복적으로 조정한다. 또한, 깊이 추정 훈련 시스템(170)은 파라미터들을 정제할 때 최소화되는, 합성 프레임과 이미지의 비교에 기초하여, 입력 스케일 폐색 및 동작 인식 손실 함수로 손실 값을 정의할 수 있다. 손실 값은 또한 깊이 추정 모델이 깊이 정보를 추정할 때 충분히 훈련되었는지 및/또는 충분히 정확한지 여부를 나타낼 수 있다. 손실 함수는 또한 다른 이미지에 시간적으로 인접한 단안 비디오로부터의 시간 이미지들로부터 투영된 두 합성 프레임들 사이의 최소 측광 재구성 오차도 고려할 수 있다. 외관 정합 손실 계산에 영향을 주게 되는 합성 프레임 생성 중에, 업샘플링된 깊이 특징들도 사용될 수 있다. 손실 함수는 손실 값을 계산할 때 이미지들 중의 정적 특징들을 고려에 넣지 않은 생성된 마스크를 구현할 수도 있다. 깊이 추정 모델이 훈련된 때, 그 깊이 추정 모델은 이미지 데이터를 수신하고, 이미지 데이터에 환경의 깊이 정보를 출력한다. 깊이 추정 훈련 시스템(170)은 훈련된 모델을 클라이언트 장치(110)에 제공한다. 깊이 추정 훈련 시스템(170)에 의한 훈련에 대해서는 아래에서 더 설명한다.

[0038] 네트워크(105)는 근거리 네트워크(예를 들어, 인트라넷), 광역 네트워크(예를 들어, 인터넷), 또는 이들의 어떤 조합과 같은 임의의 유형의 통신 네트워크일 수 있다. 네트워크는 또한 클라이언트 장치(110)와 게임 서버(120) 사이의 직접 연결을 포함할 수 있다. 일반적으로, 게임 서버(120)와 클라이언트 장치(110) 간의 통신은 다양한 통신 프로토콜(예를 들어, TCP/IP, HTTP, SMTP, FTP), 인코딩 또는 포맷(예를 들어, HTML, XML, JSON), 및/또는 보호 체계(예를 들어, VPN, 보안 HTTP, SSL)를 사용하여, 임의의 유형의 유선 및/또는 무선 연결을 사용하는 네트워크 인터페이스를 통해, 실행될 수 있다.

[0039] 본원에서 논의되는 기술은 서버, 데이터베이스, 소프트웨어 애플리케이션, 및 기타 컴퓨터 기반 시스템뿐만 아니라 이러한 시스템과 주고 받는 동작 및 정보를 참조한다. 컴퓨터 기반 시스템의 고유한 유연성은 구성 요소들 간의 작업 및 기능을 아주 다양하게 구성, 조합, 및 분할할 수 있게 한다는 것을 본 기술분야의 숙련인은 인식할 것이다. 예를 들어, 본원에 논의된 서버 프로세스는 단일 서버 또는 조합되어 작동하는 다수의 서버를 사용하여 구현될 수 있다. 데이터베이스들 및 애플리케이션들은 단일 시스템에서 구현되거나 다수의 시스템에 분산될 수 있다. 분산된 구성 요소들은 순차적으로 또는 병렬로 작동할 수 있다.

[0040] 또한, 본원에 논의된 시스템 및 방법이 사용자에 대한 개인 정보에 액세스하여 분석하거나, 또는 위치 정보와 같은 개인 정보를 사용하는 경우에는, 사용자에게 프로그램들 또는 특징들이 그러한 정보를 수집할지 여부를 제어할 수 있는 기회와, 시스템 또는 다른 애플리케이션으로부터 콘텐츠를 수신할지 그리고/또는 어떻게 수신할지 여부를 제어할 수 있는 기회를 제공할 수 있다. 사용자에게 어떤 정보가 수집되는지와 그 정보가 어떻게 사용되는지에 대한 의미 있는 통지가 제공되기 전에는 그러한 정보나 데이터는 수집되거나 사용되지 않는다. 정보는 사용자가 동의하지 않는 한 수집 또는 이용되지 않으며, 이는 사용자가 언제든지 철회하거나 수정할 수 있다. 따라서, 사용자는 사용자에 대한 정보가 애플리케이션 또는 시스템에 의해 수집되고 사용되는 방법을 제어할 수 있다. 또한, 특정 정보 또는 데이터는 저장 또는 사용되기 전에 한 가지 이상의 방식으로 처리될 수 있으므로 개인 식별 정보가 제거된다. 예를 들어, 사용자에 대한 개인 식별 정보를 확인할 수 없도록 사용자의 신원을 처리할 수 있다.

[0041] 예시적인 게임 인터페이스

[0042] 도 3은 클라이언트(120)의 디스플레이 상에 플레이어와 가상 세계(210) 사이의 인터페이스의 일부로서 제공될 수 있는 게임 인터페이스(300)의 일 실시형태를 도시하고 있다. 게임 인터페이스(300)는 가상 세계(210) 및 게임의 다양한 기타 양상들, 예컨대 가상 세계(210)에서의 플레이어 위치(222) 및 가상 요소들(230)의 위치, 가상 아이템들(232), 및 가상 에너지(250)를 디스플레이하는 데 사용될 수 있는 디스플레이 윈도우(310)를 포함한다. 사용자 인터페이스(300)는 또한 그 밖의 다른 정보, 예컨대 게임 데이터 정보, 게임 통신, 플레이어 정보, 클라이언트 위치 확인 명령, 및 게임과 연관된 기타 정보를 디스플레이할 수 있다. 예를 들어, 사용자 인터페이스는 플레이어 정보(315), 예컨대 플레이어의 이름, 경험 수준, 및 기타 정보를 디스플레이할 수 있다. 사용자 인터페이스(300)는 다양한 게임 설정 및 게임 관련 기타 정보에 액세스하기 위한 메뉴(320)를 포함할 수 있다. 사용자 인터페이스(300)는 또한 게임 시스템과 플레이어 사이의 통신 및 병렬 현실 게임의 한 명 이상의 플레이어들 사이의 통신을 가능하게 하는 통신 인터페이스(330)를 포함할 수 있다.

[0043] 본 개시내용의 양태들에 따르면, 플레이어는 현실 세계에서 클라이언트 장치(110)를 단순히 들고 다님으로써 병렬 현실 게임과 상호 작용할 수 있다. 예를 들어, 플레이어는 단순히 스마트폰에서 병렬 현실 게임과 연관된 애플리케이션에 액세스하여서 그 스마트폰을 가지고 현실 세계를 돌아다님으로써 게임을 할 수 있다. 이와 관련하여, 플레이어는 위치 기반 게임을 하기 위해 디스플레이 화면에서 가상 세계의 시각적 표현을 지속적으로 볼 필요가 없다. 그 결과, 사용자 인터페이스(300)는 사용자가 게임과 상호 작용할 수 있도록 하는 복수의 비시각적

요소를 포함할 수 있다. 예를 들어, 게임 인터페이스는 플레이어가 게임 내의 가상 요소 또는 물체에 접근하는 때에나 또는 병렬 현실 게임에서 중요한 이벤트가 발생한 때에 플레이어에게 가청 알람을 제공할 수 있다. 플레이어는 이러한 가청 알람을 오디오 제어기(340)로 제어할 수 있다. 가상 요소 또는 이벤트의 유형에 따라 다양한 유형의 가청 알람이 사용자에게 제공될 수 있다. 가청 알람은 플레이어가 가상 요소 또는 물체에 근접한 정도에 따라 빈도 또는 볼륨을 증가시키거나 줄일 수 있다. 진동 알람 또는 다른 적절한 알람 또는 신호와 같은 그 밖의 다른 비시각적 알람 및 신호가 사용자에게 제공될 수 있다.

[0044] 본 기술분야의 숙련인이 본원에 제공된 개시내용을 사용하여 이해하게 되는 바와 같이, 본 개시내용에 비추어 볼 때 수많은 게임 인터페이스 구성 및 기본 기능이 명백할 것이다. 본 개시내용은 임의의 하나의 특정 구성으로 제한되도록 한 것이 아니다.

[0045] 깊이 추정 훈련

[0046] 깊이 추정 훈련 시스템(170)은 클라이언트 장치(110)가 사용하기 위한 깊이 추정 모델(130)을 훈련시킨다. 도 1에 도시된 실시형태에서, 깊이 추정 훈련 시스템(170)은 깊이 및 포즈 모델(175), 이미지 합성 모듈(180), 오차 계산 모듈(185), 외관 정합 손실 모듈(190), 스케일링 모듈(195), 및 마스크 모듈(197)을 포함한다. 다른 실시형태에서, 깊이 추정 훈련 시스템(170)은 여러 가지 구성 요소들 및/또는 추가 구성 요소들, 예를 들어 데이터 저장소, 피드백 모듈, 평활화 모듈 등을 포함할 수 있다. 예를 들어, 데이터 저장소는 깊이 및 포즈 모델(175)을 훈련시킬 때 훈련 데이터 또는 훈련된 파라미터들을 저장할 수 있다. 다른 예에서, 평활화 모듈은 깊이 맵을, 예컨대 깊이 맵에 있어서의 평활화 깊이 값들을 처리할 수 있다. 깊이 및 포즈 모델(175)은, 이미지를 수신하고 그 이미지의 깊이 특징 및/또는 포즈를 결정할 수 있는 하나 이상의 모델을 포함한다. 아래에서 논의되는 바와 같이, 깊이 및 포즈 모델(175)에는, 포즈 모델에 대한 파라미터들과 구별되는, 깊이 모델에 대한 파라미터들이 구성될 수 있다. 대안적으로, 깊이 및 포즈 모델(175)은 그 포즈 모델로부터의 하나 이상의 파라미터가 깊이 모델과 공유되도록 구성될 수 있다.

[0047] 이제 도 4를 참조하면, 깊이 추정 훈련 시스템(170)은 장면의 깊이 맵과 입력 이미지의 포즈가 입력 이미지로부터 결정될 수 있도록 하기 위해 깊이 추정 모델(410)과 포즈 추정 모델(440)을 서로 별도로 훈련시킬 수 있다. 이 실시형태에서, 깊이 추정 모델(410)과 포즈 추정 모델(440)은 각각 작동에 필요한 연산 시간 및 자원들을 사용하여 개별적으로 작동한다.

[0048] 깊이 추정 모델(410)은 입력 이미지를 수신하여 그 입력 이미지에 대응하는 깊이 맵을 결정한다. 일 실시형태에서, 깊이 추정 모델(410)은 입력 이미지를 깊이 인코더(420)를 통해 공급하여 추상 깊이 특징들을 추출한다. 깊이 인코더(420)는 이러한 특징들을 여러 가지 기계 학습 알고리즘 및 기술을 사용하여 추출할 수 있다. 이 예시에서, 깊이 인코더(420)는 추출된 특징들의 차원을 각각의 후속 계층이 축소시키게 되는 다중 계층을 포함하는 합성곱 신경망(convolutional neural network)이다. 예를 들어, 10^6 정도의 픽셀들 또는 데이터 포인트들 상의 입력 이미지는 첫 번째 계층 다음에서는 10^5 정도의 특징들의 세트로 축소된다. 깊이 인코더(420)의 최종 계층에 의해서는 추상 깊이 특징들이 10^4 정도이거나 또는 이보다 작을 수 있다. 이들 수는 순전히 예시의 목적으로 제공되는 것이다. 실제로, 깊이 인코더는 여러 가지 수의 계층을 가질 수 있으며, 픽셀들 및 깊이 특징들의 수는 다를 수 있다.

[0049] 이와 반대로, 깊이 디코더(430)는 추상 특징들의 차원을 증가시키기 위해 다중 계층을 포함한다. 위의 예에 따라, 깊이 디코더(430)는 10^4 정도의 추상 깊이 특징들을 취할 수 있고, 입력 이미지의 각 픽셀에서의 깊이를 점차적으로 - 다중 계층에 걸쳐 - 추론할 수 있다. 그런 다음에 깊이 디코더(430)는 깊이 맵을 출력하는데, 여기서 깊이 맵 상의 각 픽셀은 그 픽셀에 투영되는 장면 내의 가장 가까운 물체의 거리에 해당한다. 대안적인 실시형태에서, 깊이 디코더(430)는 시차 맵(disparity map)을 출력하는데, 여기서 시차 맵 상의 각 픽셀은 상기 거리의 역에 해당한다. 본 개시내용 전체에 걸쳐 깊이 맵과 관련하여 설명된 원리는 시차 맵이 있는 구현에 쉽게 적용된다. 예를 들어, 입력 이미지는 카메라로부터 어떤 미지의 거리에 떨어진 나무를 소정의 픽셀로 캡처했다. 깊이 디코더(430)는 카메라로부터 그 픽셀의 블록까지의 거리에 대응하는 깊이 값을 출력한다. 일부 실시형태에서, 출력 깊이 값들은 다른 것에 상대적인 것일 수 있거나, 또는 본질적으로 정의될 수 있다. 다른 실시형태에서, 출력 깊이 값은 장면 내의 물체의 실제 측정치, 예를 들어 블록이 10 피트 떨어져 있다거나 또는 건물이 25 미터 떨어져 있다는 것과 같은, 실제 측정치를 제공하는 스케일까지 이른다.

[0050] 포즈 추정 모델(440)은 복수의 연쇄 연결(concatenate)된 입력 이미지들을 수신하여 그 입력 이미지들 각각에 대한 포즈를 결정한다. 포즈는 일반적으로 두 이미지의 관점들(perspectives) 간의 수학적 변환을 지칭한다. 본 개시내용 전체에서, 포즈는 한 이미지의 원근법을 일반적으로 설명하는데, 이 원근법은 두 이미지 사이의 변환

을 정의하는 데 사용될 수 있다. 연쇄 연결된 입력 이미지들은 그 연쇄 연결된 입력 이미지들로부터 추상 포즈 특징들을 추출하는 포즈 인코더(450)에 입력된다. 그 다음 추상 포즈 특징들이, 연쇄 연결된 입력 이미지들 각각에 대한 포즈 또는 각 쌍의 입력 이미지들 간의 상대적 변환을 결정하는 포즈 디코더(460)를 통해 입력된다. 포즈 인코더(450)는, 추상 포즈 특징들을 추출한 다음 연쇄 연결된 입력 이미지들 각각에 대한 포즈를 추론하기 위한 다중 계층을 포함하는 합성곱 신경망으로서 구성될 수 있다.

[0051] 대안적인 구성에서는, 깊이-포즈 하이브리드 모델(455)이 포즈 추정 모델로부터의 파라미터들을 깊이 모델과 공유하는데, 이는, 여러 이점들 중에서도 특히, 훈련될 파라미터가 적게 주어지면 전체 계산 시간을 감소시킨다. 일 실시형태에서, 깊이-포즈 하이브리드 모델(455)은 장면의 하나 이상의 이미지를 수신하여 그 이미지의 하나 이상의 포즈를 결정하는 모델이다. 깊이-포즈 하이브리드 모델(455)은 깊이 추정 모델(410)의 깊이 인코더(420)일 수 있는 깊이 인코더(470)와, 포즈 추정 모델(440)로부터의 포즈 디코더(460)일 수 있는 포즈 디코더(480)를 포함한다. 이 실시형태에서, 깊이-포즈 하이브리드 모델(455)은 깊이 추정 모델(410)과 포즈 추정 모델(440)에서 사용되는 원리들을 포함하고, 따라서 전체 컴퓨팅 시간 및 자원을 줄일 수 있다. 더욱이, 깊이-포즈 하이브리드 모델(455)은 두 모델 간에 정보를 공유하는 수단을 제공하여 훈련을 더 쉽게 할 수 있게 한다.

[0052] 깊이-포즈 하이브리드 모델(455)은 깊이 인코더(470)와 포즈 디코더(480)를 사용한다. 일 실시형태에서, 깊이-포즈 하이브리드 모델(455)은 복수의 입력 이미지를 취하여서 그 입력 이미지 각각을 깊이 인코더(470)를 통해 공급하여 추상 깊이 특징들을 추출한다. 각각의 입력 이미지로부터의 추상 깊이 특징들은 포즈 디코더(480)에 입력되기 전에 함께 연쇄 연결되어, 입력 이미지들 각각에 대한 포즈 또는 2개의 후속 입력 이미지들 사이의 상대 변환을 생성한다. 깊이-포즈 하이브리드 모델(455)은 각 쌍의 입력 이미지들에 대한 포즈를 추출하는 데 있어서 포즈 추정 모델(440)보다 전산적 측면에서 더 효율적이다. 깊이-포즈 하이브리드 모델(455)은 입력 이미지들을 연쇄 연결하는 포즈 추정 모델(440)과 대조적으로 일부 입력 이미지들의 추상 깊이 특징들을 연쇄 연결한다. 깊이-포즈 하이브리드 모델(455)의 포즈 디코더(480)는 깊이 인코더(470)와 그 포즈 디코더(480) 사이에 훈련 파라미터들을 공유함으로써 외부 컴퓨팅 자원의 사용을 줄일 수 있다.

[0053] 이미지 합성 모듈(180)은 한 훈련 이미지로부터 다른 훈련 이미지로 합성 프레임을 투영한다. 단안 비디오 데이터의 투영에 있어서, 이미지 합성 모듈(180)은 제1 시간 단계의 한 시간 이미지를 제2 시간 단계의 제2 시간 이미지에 투영되되, 제1 시간 이미지의 깊이와, 제1 시간 이미지 시간 단계에서 제2 시간 이미지 시간 단계로의 상대 변환을 모두 고려하여, 투영한다. 깊이는 중간 변수인 반면, 상대 변환은 깊이 및 포즈 모델(175)에서 획득된다.

[0054] 추가 실시형태에서, 이미지 합성 모듈(180)은 또한 각 이미지의 본질을 고려한다. 이미지의 본질은 해당 이미지를 캡처하는 데 사용된 카메라의 기하학적 속성, 예를 들어, 카메라의 초점 거리, 카메라의 주점 오프셋, 카메라의 비대칭도를 포함하는 카메라의 기하학적 속성을 지칭한다. 일부 경우에서, 촬영한 모든 이미지들 사이에서 각 카메라에 대해 본질이 일정할 수 있거나, 카메라가 다양한 이미지를 촬영할 때 파라미터를 조정함에 따라 본질이 달라질 수 있다. 어느 경우이든지 간에 이미지의 본질은 시간 이미지를 변환하는 데에도 사용되는 고유 행렬로서 표현될 수 있다. 추가 실시형태에서, 이미지 합성 모듈(180)은 또한 이미지의 포즈를 사용하여 이미지들을 단안 훈련 이미지 데이터로 와핑(warp)한다. 이미지 합성 모듈(180)은 제1 시간 이미지를 제2 시간 이미지의 합성 프레임으로 변환한다.

[0055] 일 실시형태에서, 이미지 합성 모듈(180)은 단안 비디오로부터 연속적인 3개의 시간 이미지들의 세트를 취하고, 제1 시간 이미지를 제2 시간 이미지 시간 단계에 제1 합성 프레임으로서 투영한다. 이미지 합성 모듈(180)은 또한 제3 시간 이미지를 제2 시간 이미지 시간 단계에 제2 합성 프레임으로서 투영한다. 입체 이미지 데이터의 투영에 있어서, 이미지 합성 모듈(180)은 입체 이미지 쌍 중 한 이미지(왼쪽 이미지)를 입체 이미지 쌍 중 다른 이미지(오른쪽 이미지)로 투영한다. 이미지 합성 모듈(180) - 단안 비디오 데이터의 투영과 유사함 - 은 입체 이미지 쌍의 깊이와, 한 이미지에서 다른 이미지로 투영할 때의 왼쪽 이미지와 오른쪽 이미지 사이의 포즈를 모두 고려한다. 그러나 단안 비디오 데이터와는 달리, 왼쪽 이미지와 오른쪽 이미지 사이의 포즈는 입체 이미지 쌍을 캡처한 두 대의 카메라의 배치에 의해 결정된다. 이미지 합성 모듈(180)은 왼쪽 이미지를 오른쪽 이미지에 오른쪽 합성 프레임으로서 투영하고, 오른쪽 이미지를 왼쪽 이미지에 왼쪽 합성 프레임으로서 투영한다.

[0056] 오차 계산 모듈(185)은 합성 프레임들과 시간 이미지들 간의 차이를 계산한다. 단일 입력 이미지로 측광 재구성 오차를 계산하는 일 실시형태에서, 오차 계산 모듈(185)은 단일 소스 이미지로부터 투영된 합성 프레임과 다른 이미지 사이의 차이를 측광 재구성 오차로 취한다.

[0057] 외관 정합 손실 모듈(190)은 측광 재구성 오차를 다수의 입력 이미지로 계산할 때 결정한다(이를 외관 정합 손

실이라고도 함). 3개의 연속하는 이미지들의 세트를 갖는 위의 실시형태에 따라, 오차 계산 모듈(185)은 제1 합성 프레임과 제2 시간 이미지 사이의 차이와, 제2 합성 프레임과 제2 시간 이미지 사이의 다른 차이를 계산할 수 있다. 하나의 시간 이미지에 있는 특징들이 인접한 시간 이미지에서 가려졌거나 또는 드러난 때 문제가 일어날 수 있다. 안타깝게도 이러한 특징들에 대응하는 픽셀들은 깊이 모델의 훈련에 부정적인 영향을 미칠 수 있다. 예를 들어, 이러한 픽셀에 대해 정확한 깊이가 예측되는 경우, 폐색된(또는 폐색에서 벗어난) 소스 이미지에 있어서의 대응하는 측광 재구성 오차는 매우 클 가능성이 있으며, 이는 깊이가 정확하게 예측됨에도 불구하고 측광 재구성 오차가 커지는 불이익을 유발한다. 이러한 문제가 있는 픽셀은 두 가지 주요 범주, 즉 이미지 경계에서의 자체 움직임(egomotion)으로 인한 시야 밖의 픽셀과, 가려진(또는 드러난) 픽셀에서 나온다. 일 실시형태에서, 외관 정합 손실 모듈(190)은 제1 합성 프레임과 제2 합성 프레임으로부터의 두 가지 차이 사이의 최소값을 식별한다. 다른 실시형태에서, 외관 정합 손실 모듈(190)은 상기 두 가지 차이를 평균한다. 입체 이미지 쌍들을 갖는 위의 실시형태에 따라, 오차 계산 모듈(185)은 왼쪽 합성 프레임과 왼쪽 이미지 사이의 차이인 왼쪽 차이와, 오른쪽 합성 프레임과 오른쪽 이미지 사이의 차이인 오른쪽 차이를 계산할 수 있다. 외관 정합 손실 모듈은 최소값을 식별하거나, 왼쪽 차이와 오른쪽 차이 사이의 평균을 계산할 수 있다. 두 가지 차이 사이에서 최소 값을 취하는 것은 가려진 물체들이 하나의 뷰에는 있지만 다른 뷰에는 없는 상태로 발생하는 문제를 완화시켜서 아티팩트가 생성되는 것을 피하는 데 도움이 된다. 이는 또한, 이미지 경계에서 아티팩트를 현저히 감소시키고, 폐색 경계의 선예도(sharpness)를 개선하고, 깊이 추정 시 전체적으로 더 우수한 정확도를 이끌어내는 데 있어서 유리하다.

[0058] 스케일링 모듈(195)은 깊이 맵을 훈련에 사용하기 위해 입력 이미지의 해상도로 스케일링한다. 통상적으로, 외관 정합 손실은 깊이 디코더의 각 계층에서의 개별 손실들의 조합으로 계산된다. 스케일링 모듈(195)은 업샘플링되는 훈련 이미지의 깊이 특징들에 대한 스케일링 인자를 깊이 특징들의 해상도와 입력 이미지의 해상도에 기초하여 결정한다. 업샘플링은 이중선형 샘플링(bilinear sampling) 또는 이중삼차 샘플링(bicubic sampling)을 포함하지만 이에 국한되지 않는 다양한 이미지 업샘플링 기술로 달성할 수 있다. 업샘플링된 깊이 특징들은 합성 프레임 생성 및 외관 정합 손실 계산에 사용된다. 업샘플링된 깊이 특징들을 사용하게 되면, 향상된 훈련 결과가 제공되며, 깊이 디코더에서의 각 계층의 해상도의 이미지들에서 측광 재구성 오차를 계산할 때 끌어 들어올 수 있는 텍스처 복사 아티팩트(즉, 입력 이미지들로부터 잘못 전달된 깊이 맵의 세부 사항)를 방지할 수 있다.

[0059] 마스킹 모듈(197)은 훈련 이미지 데이터에서 하나 이상의 정적 특징을 마스킹한다. 정적 특징은 예를 들어 단안 비디오에서 2개 이상의 이미지들 사이에서 실질적으로 유사한 위치에 있는 픽셀들의 세트로 정의될 수 있다. 예를 들어, 단안 비디오를 캡처하는 카메라와 동일한 속도로 움직이는 물체는 프레임마다 단안 비디오에서 실질적으로 유사한 위치에 픽셀로 표시될 것이다. 즉, 물체는 제1 타임 스탬프의 제1 이미지와 제2 타임 스탬프의 제2 이미지 사이에서 실질적으로 동일한 위치에 나타날 수 있다. 마스킹 모듈(197)은, 깊이 추정 훈련 시스템(170)이 외관 정합 손실을 계산할 때 정적 특징들을 필터링하는 마스크를 정적 특징들에 마스크를 적용함으로써 이러한 정적 특징들을 설명한다. 이렇게 하게 되면, 깊이 모델이 단안 비디오의 정적 특징을 심히 부정확한 깊이에 있는 것으로 판단하는 것이, 예를 들면, 무한대 쪽으로 향하는 물체는 프레임마다 정적으로 보이기 때문에 무한대를 향하는 경향이, 방지된다.

[0060] 일 구현예에서, 마스킹 모듈(197)은 계산된 손실에 기초하여 마스크를 적용한다. 마스킹 모듈(197)은 제1 시간 이미지와 제2 시간 이미지 간의 제1 손실을 계산한다. 마스킹 모듈(197)은 제1 시간 이미지와 제2 시간 이미지에서 투영된 합성 프레임 간의 제2 손실을 별도로 계산한다. 마스크는 제1 손실이 제2 손실보다 큰지 여부에 따라 크로네커 델타 함수(Kronecker delta function)일 수 있다. 그 다음 마스크는 깊이 모델의 파라미터들에 대한 훈련 중에 합성 프레임들과 입력 이미지들 사이의 손실 계산에 적용될 수 있다.

[0061] 깊이 추정 훈련 시스템(170)은 그의 모델들과 모듈들을 훈련 이미지들로 훈련시킨 후에, 깊이 추정 모델(130)이 색상 입력 이미지를 수신하여서 깊이 및 포즈 모델(175), 이미지 합성 모듈(180), 오차 계산 모듈(185), 외관 정합 손실 모듈(190), 및 스케일링 모듈(195)을 포함하는 깊이 추정 훈련 시스템(170)에 의해 훈련된 파라미터들에 기초하여 깊이 맵을 생성하도록, 파라미터들을 제공할 수 있다. 깊이 추정 훈련 시스템(170)이 편의를 위해 게임 서버(120)의 일부로 도시되었지만, 그 모델들의 일부 또는 전부는 다른 컴퓨팅 장치에 의해 훈련될 수 있으며, 또한 게임 애플리케이션에 포함되거나 또는 요구에 따라 클라우드에서 액세스되는 운영 시스템의 일부가 되는 것을 포함하여 다양한 방식으로 클라이언트 장치(110)에 제공될 수도 있다.

[0062] 도 5는 하나 이상의 실시형태에 따른 것으로, 깊이 추정 모델을 훈련하는 일반적인 프로세스(500)을 설명하는 흐름도이다. 이 프로세스(500)는 깊이 추정 모델(130)이 입력 이미지를 받은 깊이 맵을 생성할 때 사용하는 복

수의 파라미터를 생성한다.

- [0063] 깊이 추정 훈련 시스템(170)은 먼저, 복수의 단안 시간 이미지들 및/또는 복수의 입체 이미지 쌍들의 조합을 포함하는 훈련 이미지 데이터를 획득한다(단계 510). 단안 비디오 데이터가 외부 장치 상의 카메라, 예를 들어 클라이언트 장치(110) 상의 카메라 어셈블리(125)로부터 수신될 수 있다. 입체 이미지 쌍들이 외부 장치 상의 한 쌍의 쌍안 카메라들로부터, 예를 들어 클라이언트 장치(110) 상의 카메라 어셈블리(125)로부터 수신될 수 있다. 일 실시형태에서, 네트워크 인터페이스(105)가 훈련 이미지 데이터를 수신한다. 깊이 추정 훈련 시스템(170)은 훈련 이미지 데이터를 다양한 데이터 저장소에 저장할 수 있는데, 예를 들면, 단안 비디오 데이터를 단안 비디오 데이터 저장소에 저장하고 입체 이미지 쌍들을 입체 이미지 데이터 저장소에 저장할 수 있다.
- [0064] 깊이 추정 훈련 시스템(170)은 단안 비디오를 사용할 때에는 단안 비디오 데이터로부터의 시간 이미지들을 3개의 연속하는 시간 이미지들의 세트들로 그룹화한다(단계 520). 3개의 이미지들의 세트들로 그룹화하는 이 단계(520)는 제3 시간 이미지에 투영된 2개의 시간 이미지를 사용하여 측광 재구성 오차를 계산하는 것을 목표로 한다. 다른 실시형태에서, 깊이 추정 시스템(170)은 시간 이미지들을 4개의 이미지들의 세트들 또는 5개의 이미지들의 세트들 등으로 그룹화할 수 있다.
- [0065] 깊이 추정 훈련 시스템(170)은 각 이미지를 깊이 모델에 입력하여(단계 530) 깊이 특징들을 추출할 수 있도록 한다. 일 실시형태에서, 이미지들은 깊이 특징들을, 예를 들어 그 이미지들의 해상도에 있는 깊이 특징들을, 깊이 맵으로서 추출하는 깊이 추정 모델(예를 들어, 깊이 추정 모델(410))에 입력된다.
- [0066] 깊이 추정 훈련 시스템(170)은 각 이미지에 대한 포즈를 추출하는 포즈 디코더에 이미지들을 입력한다(단계 540). 일 실시형태에서, 이미지들은 이미지에 대한 포즈를 추출하는 포즈 추정 모델(예를 들어, 포즈 추정 모델(440))에 입력된다. 깊이-포즈 하이브리드 모델을 갖는 실시형태에서, 깊이 인코더(예를 들어, 깊이 인코더(470))로부터 결정된 추상 깊이 특징들이 연쇄 연결되고 포즈 디코더(예를 들어, 포즈 디코더(480))에 입력되어 각 시간 이미지에 대한 포즈를 추출할 수 있도록 한다. 입체 이미지 쌍들을 사용하게 되면, 포즈는 입체 이미지 쌍의 두 관점들 간의 변환을 정의하거나, 또는 그 변환을 정의하는 데 도움이 된다. 일부 실시형태에서, 입체 이미지 쌍의 두 관점들 사이의 포즈는 정해진 것이고/것이거나 알려진 것이다. 깊이 추정 훈련 시스템(170)은 3개의 연속하는 시간 이미지들(예를 들어, 제1 시간 이미지, 제2 시간 이미지, 및 제3 시간 이미지)의 세트들로 그룹화된 단안 비디오 데이터를 사용하여, 제1 이미지에서 제2 이미지로의 상대 변환과, 제2 이미지에서 제3 이미지로의 또 다른 상대 변환을 추출한다.
- [0067] 깊이 추정 훈련 시스템(170)은 깊이 특징들 및 포즈들을 사용하여, 시간 이미지들을 후속한 시간 이미지들에 그리고/또는 각각의 입체 이미지를 입체 이미지 쌍의 다른 입체 이미지에 투영한다(단계 550). 깊이 추정 훈련 시스템(170)은 각 세트의 3개의 시간 이미지들을 사용하여, 제1 시간 이미지를 제2 시간 단계에 제1 합성 프레임으로서 투영하고 제3 시간 이미지를 제2 시간 단계에 제2 합성 프레임으로서 투영한다. 깊이 추정 훈련 시스템(170)은 제1 시간 이미지의 포즈에 기초하거나 또는 중간 변수가 제1 시간 이미지의 깊이인, 제1 시간 이미지로부터 제2 시간 이미지로의 상대 변환에 기초하여 제1 시간 이미지를 제2 시간 단계에 투영한다. 깊이 추정 훈련 시스템(170)은 제3 시간 이미지를 제2 시간 단계에 투영하는데, 위와 마찬가지로, 여기서도 또한 중간 변수가 제3 시간 이미지의 깊이인, 제2 시간 이미지로부터 제3 시간 이미지로의 역 상대 변환을 사용하여 투영한다. 일 실시형태에서, 이미지 합성 모듈(180)은 하나의 시간 이미지로부터 합성 프레임으로서의 투영을 수행한다. 깊이 추정 훈련 시스템(170)은 입체 이미지 쌍들을 사용하여, 입체 이미지 쌍의 왼쪽 이미지를 입체 이미지 쌍의 오른쪽 이미지에 오른쪽 합성 프레임으로서 투영하고, 이와 유사하게 오른쪽 이미지를 왼쪽 이미지에 왼쪽 합성 프레임으로서 투영한다. 일 실시형태에서, 이미지 합성 모듈(180)은 왼쪽 이미지로부터 오른쪽 이미지로 또는 그 반대로의 투영을 수행한다.
- [0068] 깊이 추정 훈련 시스템(170)은 상기 합성 프레임들과 상기 이미지들의 비교에 기초하여 입력 스케일 폐색 및 동작 인식 손실 함수를 사용하여 손실 값을 계산한다(단계 560). 입력 스케일 폐색 및 동작 인식 손실 함수는 깊이 모델을 훈련하는 데 사용하기 위한 손실 값을 계산한다. 손실 함수에는 합성 프레임과 입력 이미지 사이의 픽셀마다의 측광 재구성 오차의 계산이 포함된다. 손실 함수는 또한, 외관 정합 손실 모듈(190)에서 전술한 바와 같이, 다른 이미지에 시간적으로 인접한 단안 비디오로부터의 시간 이미지들로부터 투영된 두 합성 프레임들 사이의 최소 측광 재구성 오차도 고려할 수 있다. 외관 정합 손실 계산에 영향을 주게 되는 합성 프레임 생성 중에, (예를 들어, 스케일링 모듈(195)에 의한) 업샘플링된 깊이 특징들도 사용될 수 있다. 손실 함수는 손실 값을 계산할 때 정적 특징들을 고려에 넣지 않은 마스킹 모듈(197)에 의해 생성된 마스크를 구현할 수도 있다.
- [0069] 깊이 추정 훈련 시스템(170)은 픽셀마다 측광 재구성 오차를 최소화함으로써 깊이 모델을 훈련시킨다(단계

570). 3개의 시간 이미지들의 세트들에 대해, 깊이 추정 훈련 시스템(170)은 제2 시간 이미지로부터의 제1 합성 프레임과 제2 합성 프레임의 차이에 기초하여 픽셀마다 최소 측광 재구성 오차를 식별한다. 추가 실시형태에서, 깊이 추정 훈련 시스템(170)은 합성 프레임들과 및 이미지들에 기초하여 깊이 추정 모델에 대한 전체 오차를 정의할 수 있다. 전체 오차는 예를 들어 한 쌍의 이미지들에 대한 측광 재구성 오차의 평균, 다수의 또는 모든 입력 이미지들에 대한 측광 재구성 오차의 평균 등으로 정의될 수 있다. 측광 재구성 오차(또는 전체 오차)를 최소화함에 있어서, 깊이 추정 훈련 시스템(170)은 깊이 모델에 대한 파라미터들을 정제한다. 포즈 모델의 파라미터들은 또한 측광 재구성 오차를 최소화하는 부분으로서 정제될 수도 있다. 일 실시형태에서, 깊이 추정 훈련 시스템(170)은 측광 재구성 오차를 2개의 차이들 사이의 절대 최소값으로서 계산한다. 일 실시형태에서, 외관 정합 손실 모듈(190)은 이미지 합성 모듈(180)과 협력하여 측광 재구성 오차를 최소화한다. 다른 실시형태에서, 스케일링 모듈(195)은 깊이 모델에서 각 계층의 파라미터들을 조정하기 위해 다양한 해상도의 이미지들의 깊이 맵들을 스케일링한다. 다른 실시형태에서, 마스킹 모듈(197)은 정적 특징을 갖는 하나 이상의 영역을 식별하고, 이들 영역을 측광 재구성 오차 계산 시에 마스킹한다.

[0070] 깊이 추정 모델

[0071] 도 6은 하나 이상의 실시형태에 따른 것으로, 깊이 추정 모델을 사용하는 일반적인 프로세스(600)를 설명하는 흐름도이다. 이 프로세스(600)는 입력 이미지를 받은 깊이 맵을 생성한다. 이 프로세스(600)는 훈련된 깊이 추정 모델을 갖춘 클라이언트 장치에 의해 달성될 수 있다. 클라이언트 장치는 일반 컴퓨팅 장치일 수 있으며, 카메라도 구비할 수 있다. 일부 실시형태에서, 클라이언트 장치는 위의 도 1 내지 도 3에서 설명된 병렬 현실 게임에서 구현된다. 다음 설명은 클라이언트 장치의 맥락 내에 있지만, 이 프로세스(600)는 다른 컴퓨팅 장치에서 수행될 수 있다.

[0072] 본 방법은 장면의 이미지를 수신하는 단계(610)를 포함한다. 장면의 이미지는, 클라이언트 장치의 구성 요소이거나 또는 클라이언트 장치 외부에 있는 카메라에 의해 캡처될 수 있다. 병렬 현실 게임과 관련하여, 장면은 가상 세계의 가상 위치에 매핑될 수 있는 실제 위치일 수 있다. 장면의 이미지는 이미지를 캡처한 카메라의 기하학적 속성에 해당하는 본질도 가질 수 있다. 이미지는 카메라가 캡처한 단일 이미지일 수 있다. 대안적으로, 이미지는 카메라에 의해 캡처되는 비디오의 프레임일 수 있다.

[0073] 본 방법은 장면의 이미지를 훈련된 깊이 추정 모델에 입력하는 단계(620)를 포함한다. 깊이 추정 모델은 깊이 추정 시스템(170)에 의해 예를 들어 도 5의 프로세스(500)를 통해 훈련될 수 있다. 깊이 추정 모델은 장면의 이미지를 수신하고, 선택적으로, 이미지의 본질도 수신한다.

[0074] 본 방법은 훈련된 깊이 추정 모델에 의해 장면의 이미지에 대응하는 장면의 깊이 맵을 생성하는 단계(630)를 포함한다. 깊이 맵의 각 픽셀은 장면 이미지의 대응하는 픽셀에서의 표면의 상대 거리를 설명하는 깊이 값을 갖는다. 깊이 추정은 장면의 이미지를 수신하고, 도 5에 따라 훈련된 파라미터들에 기초하여 깊이 맵을 출력한다.

[0075] 본 방법은 장면의 깊이 맵에 기초하여 가상 콘텐츠를 생성하는 단계(640)를 포함한다. 가상 콘텐츠는 예를 들어 게임 데이터베이스(115)에 저장된 병렬 현실 게임용 콘텐츠로부터 공급될 수 있다. 생성된 가상 콘텐츠는 장면의 이미지에 증강될 수 있는 증강 현실 콘텐츠일 수 있다. 예를 들어, 장면의 깊이를 이해하면서 장면 주위를 돌아다닐 수 있는 가상 캐릭터가 생성된다. 한 예에서, 가상 캐릭터는 그 가상 캐릭터가 사용자를 향해 거리를 걸어감에 따라 크기가 커질 수 있다. 다른 예에서, 가상 캐릭터는 그 가상 캐릭터의 일부를 가리는 나무 뒤에 숨어 있을 수 있다.

[0076] 본 방법은 가상 콘텐츠로 증강된 장면의 이미지를 디스플레이하는 단계(650)를 포함한다. 클라이언트 장치는 전자 디스플레이를 포함한다. 전자 디스플레이는 카메라로 캡처한 비디오의 지속적인 공급을 증강된 가상 콘텐츠와 함께 제공할 수 있다.

[0077] 위의 예에 따라, 병렬 현실 게임은 가상 캐릭터와의 상호 작용을 목표로 제공할 수 있다. 가상 캐릭터와 상호 작용하기 위해서는 모바일 장치의 사용자는 카메라의 시야에 가상 캐릭터를 유지하면서 모바일 장치를 여기저기로 움직일 필요가 있을 수 있다. 사용자가 모바일 장치를 여기저기로 움직일 때, 모바일 장치는 사용자의 모바일 장치 움직임에 따라 장면이 변경됨에 따라 장면의 깊이 정보를 반복적으로 생성하는 데 사용할 수 있는 비디오 또는 이미지 데이터를 지속적으로 캡처할 수 있다. 모바일 장치는 디스플레이 상에 비디오 피드를 업데이트할 수 있으며 또한 이와 동시에 생성된 깊이 정보에 기초하여 가상 캐릭터를 업데이트할 수 있으므로, 사용자는 가상 캐릭터를 장면 내에서 항상 적절하게 상호 작용하고 있다는 것으로 인식할 수 있는데, 그 상호 작용의 예는 물체를 자세히 보여주지 않는 것, 물체가 그의 어떤 부분도 가리지 않게 잘린 부분이 없이 보이는 것 등이

있다.

[0078] 예시적인 컴퓨팅 시스템

[0079] 도 7은 일 실시형태에 따른 컴퓨팅 장치의 예시적인 아키텍처이다. 도 7은 일 실시예에 따른 것으로, 본원에서 설명된 하나 이상의 엔티티의 일부 또는 전부로서 사용되는 컴퓨터의 물리적 구성 요소들을 예시하는 고수준의 블록도를 도시하고 있지만, 컴퓨터는 도 7에 제공된 것에 추가한 구성 요소들을 갖거나, 그보다 적게 구성된 구성 요소들을 갖거나, 변형된 구성 요소들을 가질 수 있다. 도 7은 컴퓨터(700)를 도시하고 있지만, 이 도면은 본원에 설명된 구현 예들의 구조적 개략도로써보다는 컴퓨터 시스템에 존재할 수 있는 다양한 특징들에 대한 기능 설명으로서 의도되어 있다. 실제로, 그리고 본 기술분야의 통상의 지식을 가진 자가 인식하는 바와 같이, 개별적으로 도시된 물품들은 결합될 수 있고, 일부 물품들은 분리될 수 있다.

[0080] 도 7에는 칩셋(704)에 연결된 적어도 하나의 프로세서(702)가 예시되어 있다. 또한 칩셋(704)에는 메모리(706), 저장 장치(708), 키보드(710), 그래픽 어댑터(712), 포인팅 장치(714), 및 네트워크 어댑터(716)가 연결된다. 디스플레이(718)는 그래픽 어댑터(712)에 연결된다. 일 실시형태에서, 칩셋(704)의 기능은 메모리 컨트롤러 허브(720) 및 I/O 허브(722)에 의해 제공된다. 다른 실시형태에서, 메모리(706)는 칩셋(704) 대신에 프로세서(702)에 직접 연결된다. 일부 실시형태에서, 컴퓨터(700)는 상기 구성 요소들을 상호 연결하기 위한 하나 이상의 통신 버스를 포함한다. 선택적으로, 하나 이상의 통신 버스는 시스템 구성 요소들 간의 통신을 상호 연결하고 제어하는 회로(칩셋이라고도 함)를 포함한다.

[0081] 저장 장치(708)는 임의의 비일시적 컴퓨터 판독 가능 저장 매체, 예컨대 하드 드라이브, CD-ROM(compact disk read-only memory), DVD, 또는 반도체 메모리 장치 또는 기타 광학 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 자기 디스크 저장 장치, 광 디스크 저장 장치, 플래시 메모리 장치, 또는 기타 비휘발성 반도체 저장 장치 등이다. 이러한 저장 장치(708)는 또한 영구 메모리로 지칭될 수 있다. 포인팅 장치(714)는 마우스, 트랙볼, 또는 기타 유형의 포인팅 장치일 수 있으며, 컴퓨터(700)에 데이터를 입력하기 위해 키보드(710)와 결합하여 사용된다. 그래픽 어댑터(712)는 디스플레이(718)에 이미지 및 기타 정보를 디스플레이한다. 네트워크 어댑터(716)는 컴퓨터(700)를 근거리 또는 광역 네트워크에 연결한다.

[0082] 메모리(706)는 프로세서(702)에 의해 사용되는 명령 및 데이터를 보관한다. 메모리(706)는 비영구 메모리일 수 있으며, 그 예는 DRAM, SRAM, DDR RAM, ROM, EEPROM, 플래시 메모리 등의 고속 랜덤 액세스 메모리를 포함한다.

[0083] 본 기술분야에 알려진 바와 같이, 컴퓨터(700)는 도 13에 도시된 것과 상이하고/하거나 그 밖의 다른 구성 요소들을 구비할 수 있다. 또한, 컴퓨터(700)에는 특정의 예시된 구성 요소가 없을 수 있다. 일 실시형태에서, 서버 역할을 하는 컴퓨터(700)에는 키보드(710), 포인팅 장치(714), 그래픽 어댑터(712), 및/또는 디스플레이(718)가 없을 수 있다. 또한, 저장 장치(708)는 컴퓨터(700)로부터 로컬 및/또는 원격에 있을 수 있다(예컨대, 스토리지 영역 네트워크(SAN) 내에 구현될 수 있다).

[0084] 본 기술분야에 알려진 바와 같이, 컴퓨터(700)는 본원에 설명된 기능을 제공하기 위한 컴퓨터 프로그램 모듈을 실행하도록 구성된다. 본원에서 사용되는 용어 "모듈"은 특정한 기능을 제공하기 위해 사용되는 컴퓨터 프로그램 로직을 지칭한다. 따라서, 모듈은 하드웨어, 펌웨어, 및/또는 소프트웨어로 구현될 수 있다. 일 실시형태에서, 프로그램 모듈들이 저장 장치(708)에 저장되고, 메모리(706)에 탑재되고, 프로세서(302)에 의해 실행된다.

[0085] 추가 고려 사항

[0086] 실시형태들에 대한 추가 논의는 전체 내용이 본원에 인용되어 포함되는 "자가 감독 단안 깊이 추정"에 대한 철저한 탐구"라는 제목의 명세서 부록에서 찾아볼 수 있다.

[0087] 위 설명의 일부 부분은 실시형태들을 알고리즘 프로세스 또는 작동의 관점에서 설명하고 있다. 이러한 알고리즘에 대한 서술 및 설명은 데이터 처리 기술분야의 숙련인들이 작업의 내용을 동 기술분야의 다른 숙련인에게 효과적으로 전달하는 데 일반적으로 사용하는 것이다. 이러한 작동은 기능적으로, 전산적으로, 또는 논리적으로 설명되지만, 프로세서 또는 등가 전기 회로, 마이크로코드 등에 의한 실행을 위한 명령을 포함하는 컴퓨터 프로그램에 의해 구현되는 것으로 이해된다. 또한, 이러한 기능적 작동들의 배열을 일반성을 잃지 않고 모듈이라고 지칭하는 것이 때로는 편리하다는 것도 입증되었다.

[0088] 본원에 사용된 바와 같이, "일 실시형태" 또는 "실시형태"라는 언급은 그 실시형태와 관련하여 설명된 특정 요소, 특징, 구조, 또는 특성이 적어도 하나의 실시형태에 포함된다는 것을 의미한다. 명세서의 여러 위치에서 출

현하는 "일 실시형태에서"라는 문구는 반드시 모두가 동일한 실시형태를 지칭하는 것은 아니다.

[0089] 일부 실시형태는 "결합된" 및 "연결된"이라는 표현을 이들의 파생어와 함께 사용하여 설명될 수 있다. 이러한 용어들은 서로 동의어로 의도된 것이 아님을 이해해야 한다. 예를 들어, 일부 실시형태는 둘 이상의 요소들이 서로 직접적으로 물리적 또는 전기적 접촉을 하고 있음을 나타내기 위해 "연결된"이라는 용어를 사용하여 설명될 수 있다. 다른 예로서, 일부 실시형태는 둘 이상의 요소들이 직접적으로 물리적 또는 전기적 접촉을 하고 있음을 나타내기 위해 "결합된"이라는 용어를 사용하여 설명될 수 있다. 그러나, "결합된"이라는 용어는 또한 둘 이상의 요소들이 서로 직접적으로 접촉하지 않지만 여전히 서로 협력하거나 상호 작용한다는 것을 의미할 수도 있다. 이러한 맥락에서 실시형태는 제한되지 않는다.

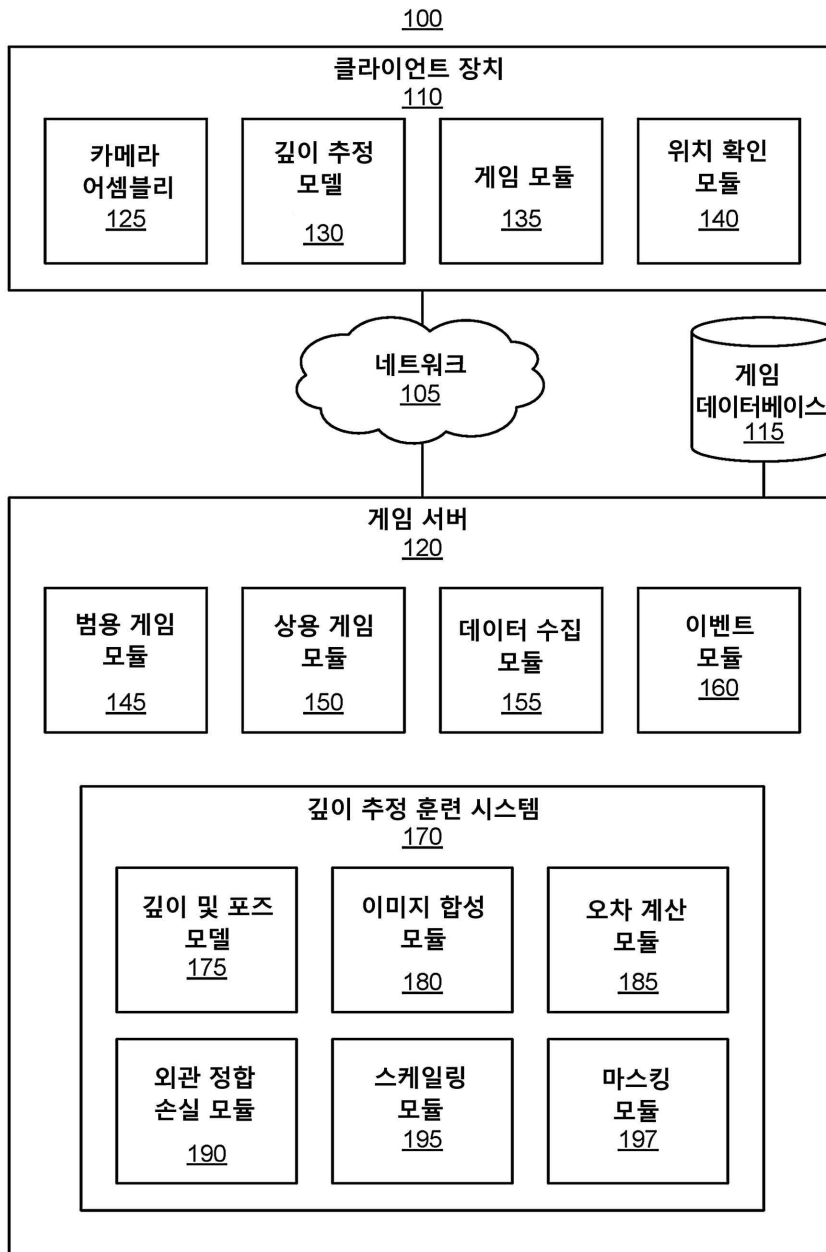
[0090] 본원에 사용된 용어 "포함하다", "포함하는", "포함한다", "포함하는", "갖는다", "갖는" 또는 이들의 임의의 다른 변형은 배타적 포함까지 망라하도록 의도된 것이다. 예를 들어, 소정의 목록의 요소들을 포함하는 프로세스, 방법, 물품, 또는 장치는 반드시 그러한 요소들로만 제한되는 것이 아니고, 명시적으로 나열되지 않았거나 그러한 프로세스, 방법, 물품, 또는 장치에 고유하지 않은 그 밖의 다른 요소들을 포함할 수 있다. 또한, 상반된 언급을 명시적으로 하지 않는 한, "또는"은 포괄적인 '또는'을 지칭하는 것이지, 배타적인 '또는'을 지칭하는 것이 아니다. 예를 들어, 조건 A 또는 B는 다음 중 어느 하나에 의해 충족된다: A는 참(또는 존재)이고 B는 거짓(또는 부재)인 것; A는 거짓(또는 부재)이고 B는 참(또는 존재)인 것; A와 B가 모두 참(또는 존재)인 것.

[0091] 또한, 부정관사("a" 또는 "an")의 용법은 실시형태의 요소들 및 구성 요소들을 설명하는 데 사용된다. 이는 단지 편의를 위해 사용하는 것이며 개시내용의 일반적인 의미를 제공하기 위해 사용하는 것이다. 이러한 설명은 하나 또는 적어도 하나를 포함하는 것으로 해석해야 하며, 단수는 달리 의미하는 것이 명백하지 않은 한 복수도 포함한다.

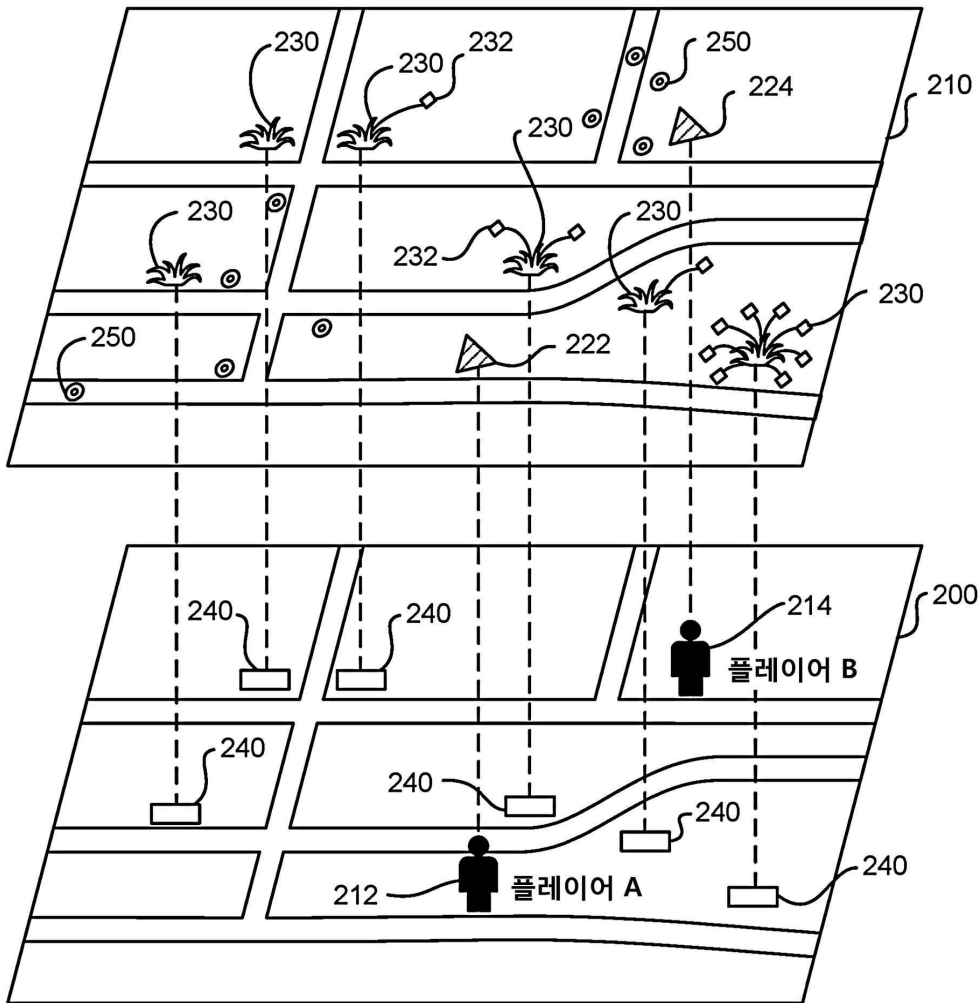
[0092] 본 기술분야의 숙련인은 본 개시내용을 읽을 때, 온라인 서비스 제공자와의 계정을 검증하는 시스템 및 프로세스를 위한 또 다른 추가적 대안의 구조적 및 기능적 설계가 진정한 비즈니스에 해당한다는 것을 이해할 것이다. 따라서, 특정 실시형태들과 적용들이 예시되고 설명되었지만, 설명된 기술적 요지는 여기에 설명된 정확한 구성 및 구성 요소로 제한되지 않는다는 것을 이해해야 하며, 그리고 본 기술분야의 숙련인에게 명백할 다양한 수정, 변경, 및 변형이 개시된 방법 및 장치의 배열, 작동, 및 세부 사항 내에서 이루어질 수 있다는 것을 이해해야 한다. 보호 범위는 다음의 청구범위에 의해서만 제한되어야 한다.

도면

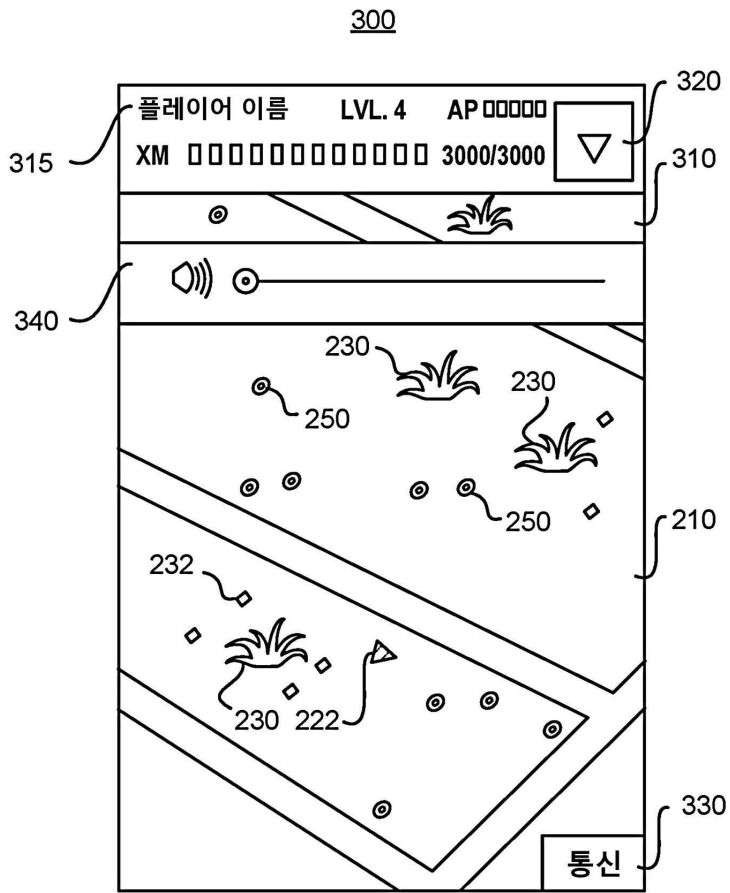
도면1



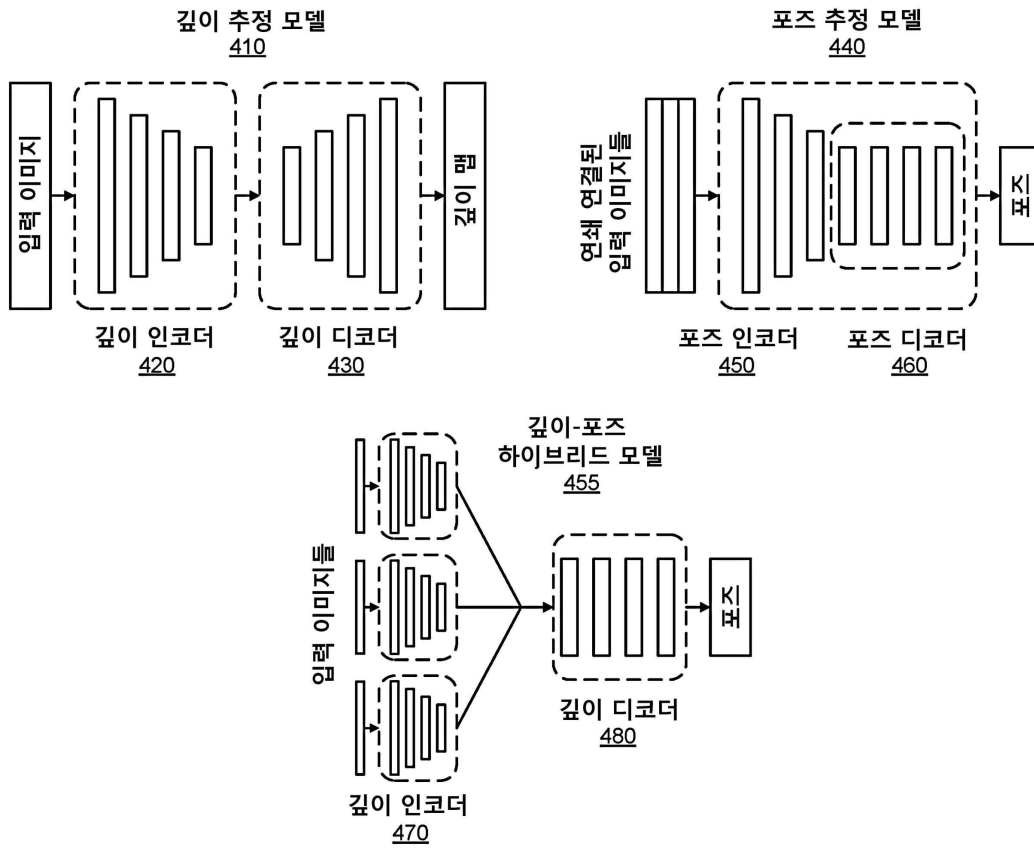
도면2



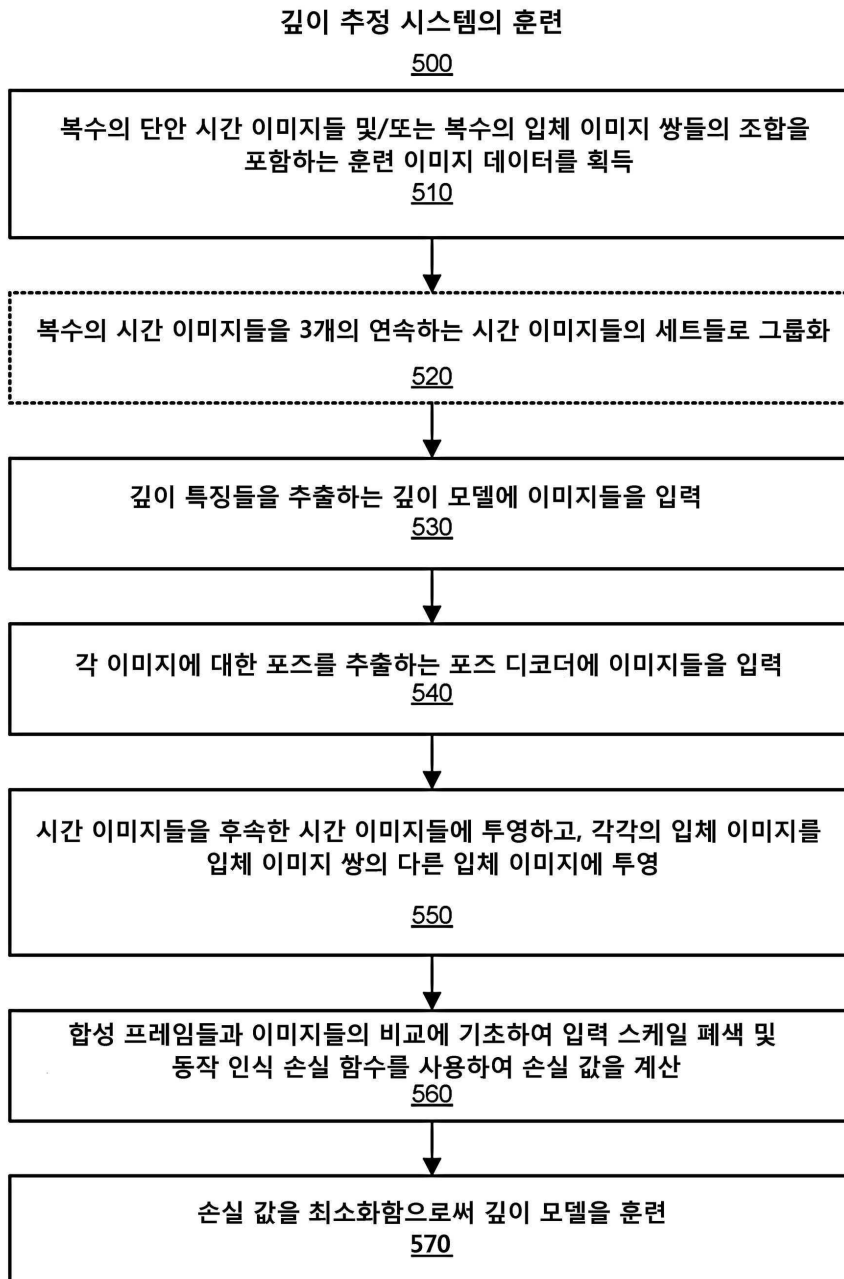
도면3



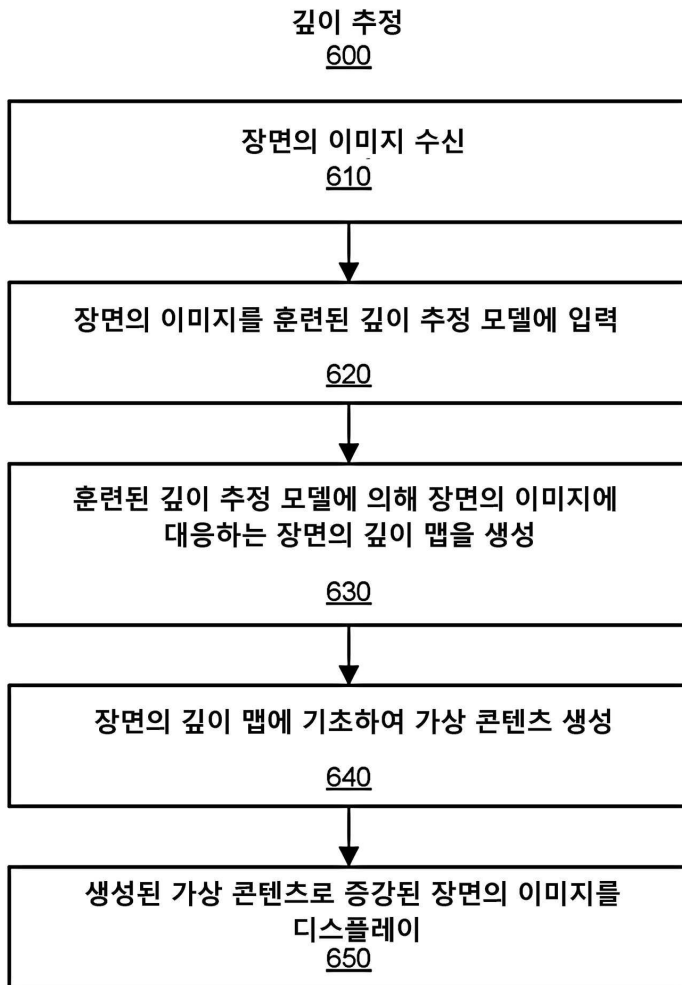
도면4



도면5



도면6



도면7

