



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0024494
(43) 공개일자 2022년03월03일

(51) 국제특허분류(Int. Cl.)
G06T 7/50 (2017.01) G06N 3/04 (2006.01)
G06N 3/08 (2006.01) G06T 7/73 (2017.01)
(52) CPC특허분류
G06T 7/50 (2017.01)
G06N 3/0454 (2013.01)
(21) 출원번호 10-2022-7000964
(22) 출원일자(국제) 2020년03월27일
심사청구일자 없음
(85) 번역문제출일자 2022년01월11일
(86) 국제출원번호 PCT/IB2020/052936
(87) 국제공개번호 WO 2020/250046
국제공개일자 2020년12월17일
(30) 우선권주장
3,046,612 2019년06월14일 캐나다(CA)

(71) 출원인
힌지 헬스 인크.
미국 94105 캘리포니아주 샌프란시스코 스위트
700 마켓 스트리트 455
(72) 발명자
브라운 콜린
캐나다 퀘벡주 에이치4제트 2제트6 몬트리올 스위트
비113 르느아르 1001
하버 루이스
캐나다 퀘벡주 에이치4제트 2제트6 몬트리올 스위트
비113 르느아르 1001
(74) 대리인
김태홍, 김진희

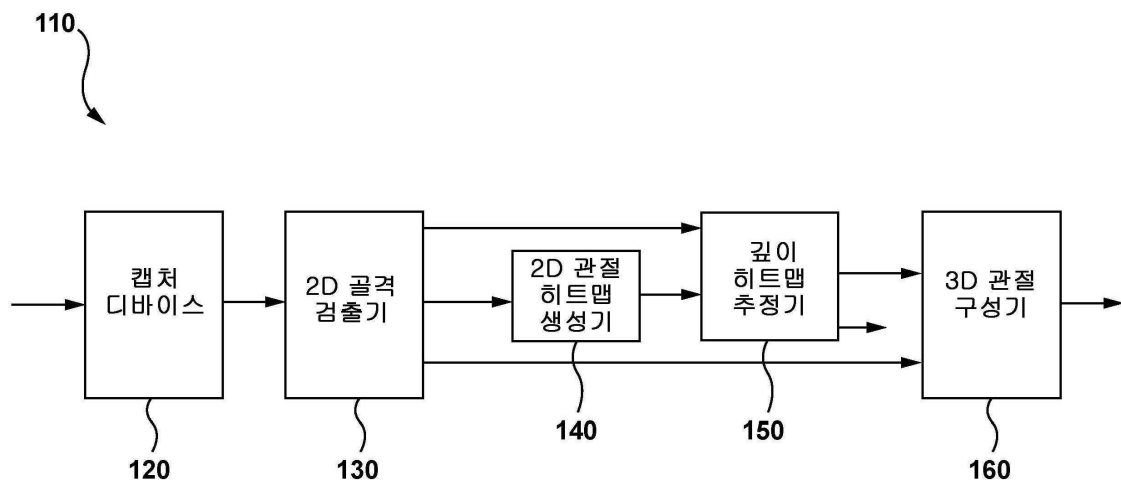
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 사람의 단안 깊이 추정을 위한 방법 및 시스템

(57) 요약

물체의 이미지 세그먼트 및 이미지 세그먼트 상의 골격 관절의 2D 위치를 포함하는 2D 관절 히트맵으로부터 골격 관절의 3D 관절 위치를 추정하기 위한 시스템 및 방법이 제공된다. 이것은 이미지 세그먼트 및 2D 관절 히트맵을, 적어도 하나의 3D 컨볼루션 계층 블록을 포함하는 컨볼루션 신경망에 적용하는 것을 포함하며, 여기서 2D 해상도는 각 3D 컨볼루션 계층에서 감소되고 깊이 해상도는 확장되어 각 관절에 대해 추정된 깊이를 생성한다. 각 관절 종류의 2D 위치를 관절 종류의 추정된 깊이와 결합하는 것은 골격 관절의 추정된 3D 관절 위치를 생성한다.

대표도



(52) CPC특허분류

G06N 3/0481 (2013.01)

G06N 3/084 (2013.01)

G06T 7/75 (2017.01)

G06T 2207/20084 (2013.01)

명세서

청구범위

청구항 1

물체의 이미지 세그먼트 및 상기 이미지 세그먼트 상의 골격 관절의 2D 위치를 포함하는 2D 관절 히트맵으로부터 골격 관절의 3D 관절 위치를 추정하기 위한 방법에 있어서,

적어도 하나의 3D 컨볼루션 계층 블록(convolutional layer block)을 포함하는 컨볼루션 신경망에 상기 이미지 세그먼트 및 상기 2D 관절 히트맵을 적용하는 단계 - 2D 해상도는 각 3D 컨볼루션 계층에서 감소되고, 깊이 해상도는 확장되어 각 관절에 대한 추정된 깊이를 생성함 -;

각 종류의 관절의 2D 위치를 상기 종류의 관절의 추정된 깊이와 결합하여 상기 골격 관절의 추정된 3D 관절 위치를 생성하는 단계

를 포함하는, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 2

제1항에 있어서,

상기 이미지 세그먼트 및 상기 2D 관절 히트맵을 상기 적어도 하나의 3D 컨볼루션 계층 이전에 적어도 하나의 2D 컨볼루션 신경망 블록에 적용하는 단계

를 더 포함하는, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 3

제1항 또는 제2항에 있어서,

상기 적어도 하나의 3D 컨볼루션 계층으로부터 깊이들을 수신하는 최대 풀링 계층(max pooling layer)을 더 포함하고, 상기 최대 풀링 계층은 상기 2D 해상도를 1로 감소시키고, 각 종류의 관절의 2D 위치를 상기 종류의 관절의 추정된 깊이와 결합하기 전에 각 관절에 대한 확장된 깊이 차원만을 남기는 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 3D 컨볼루션 계층 블록 각각은 제1 3D 컨볼루션, 제1 ReLU, 제2 3D 컨볼루션, 제2 ReLU, 최대 풀링 계층, 및 재성형 계층을 포함하는 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 5

제4항에 있어서,

각각의 3D 컨볼루션은 3x3x3보다 크지 않은 커널(kernel)을 사용하는 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 6

제4항 또는 제5항에 있어서,

상기 3D 컨볼루션 계층 블록의 최대 풀링 계층은 입력 공간 차원의 크기를 몇 배만큼 감소시키지만 상기 깊이 차원은 유지하며, 상기 재성형 계층은 상기 깊이 차원의 크기를 상기 몇 배만큼 증가시키는 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서,

상기 컨볼루션 신경망은 물체의 이미지 세그먼트 및 3D 관절 위치를 사용하는 역전파(back propagation)를 사용하여 트레이닝되는 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서,

상기 물체는 사람인 것인, 골격 관절의 3D 관절 위치를 추정하기 위한 방법.

청구항 9

2D 이미지로부터 골격 상의 관절의 3D 위치를 결정하기 위한 시스템에 있어서,

2D 이미지 및 2D 관절 골격에 대한 이미지 세그먼트로부터 하나 이상의 2D 관절 골격을 식별하는 2D 골격 검출기;

상기 하나 이상의 2D 관절 골격의 각 관절에 대한 2D 관절 히트맵을 생성하는 2D 관절 히트맵 생성기;

각 관절의 깊이를 추정하도록 구성되는 적어도 하나의 3D 컨볼루션 계층 블록을 포함하는, 상기 2D 관절 히트맵 및 상기 이미지 세그먼트를 수신하는 깊이 히트맵 추정기;

각 관절의 깊이를 상기 2D 관절 골격과 결합하여 3D 관절 골격을 구성하는 3D 관절 구성기

를 포함하는, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 10

제9항에 있어서,

상기 2D 관절 히트맵 생성기는 상기 2D 골격 검출기로부터의 2D 관절 히트맵을 사용하는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 11

제9항 또는 제10항에 있어서,

상기 깊이 히트맵 추정기는 상기 적어도 하나의 3D 컨볼루션 계층 블록 이전에 적어도 하나의 2D 컨볼루션 신경망 블록을 더 포함하는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 12

제9항 또는 제11항에 있어서,

상기 깊이 히트맵 추정기는 상기 적어도 하나의 3D 컨볼루션 계층으로부터 깊이들을 수신하는 최대 풀링 계층을 더 포함하고, 상기 최대 풀링 계층은 2D 해상도를 1로 감소시키고, 각 종류의 관절의 2D 위치를 상기 종류의 관절의 추정된 깊이와 결합하기 전에 각 관절에 대한 확장된 깊이 차원만을 남기는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 13

제9항 내지 제12항 중 어느 한 항에 있어서,

상기 3D 컨볼루션 계층 블록 각각은 제1 3D 컨볼루션, 제1 ReLU, 제2 3D 컨볼루션, 제2 ReLU, 최대 풀링 계층, 및 재성형 계층을 포함하는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 14

제13항에 있어서,

각각의 3D 컨볼루션은 3x3x3보다 크지 않은 커널을 사용하는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 15

제13항 또는 제14항에 있어서,

상기 3D 컨볼루션 계층 블록의 최대 풀링 계층은 입력 공간 차원의 크기를 몇 배만큼 감소시키지만 상기 깊이 차원은 유지하며, 상기 재구성 계층은 상기 깊이 차원의 크기를 상기 몇 배만큼 증가시키는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

청구항 16

제9항 내지 제15항 중 어느 한 항에 있어서,

상기 컨볼루션 신경망은 물체의 이미지 세그먼트 및 3D 관절 위치를 사용하는 역전파를 사용하여 트레이닝되는 것인, 관절의 3D 위치를 결정하기 위한 시스템.

발명의 설명

기술 분야

[0001] 본 개시는 단안 카메라로부터의 이미지 또는 비디오 데이터에서 사람의 관절(joints)의 상대적인 3D 위치의 추정을 위한 방법 및 시스템에 관한 것이다. 특히, 본 개시는 카메라로부터 한 명 이상의 사람의 관절의 상대적 깊이 값을 추정하는 머신 러닝 기반 접근 방식에 관한 것이다.

배경 기술

[0002] 사람 관절, 또는 예를 들어, 팔꿈치, 눈, 발가락 등과 같은 임의의 해부학적 랜드마크(landmarks)의 상대적인 3D 위치 또는 깊이 값, 즉, 카메라로부터의 거리의 단안 추정은 모션 캡처(motion capture), 가상 및 증강 현실, 스포츠 분석, 개인 훈련 및 활동 인식을 포함한 광범위한 애플리케이션을 가지고 있다.

[0003] 특히, 추가 깊이 감지 하드웨어 없이 표준 RGB 이미지 데이터로부터의 3D 관절 추정은 가장 광범위한 사용 사례 세트를 용이하게 한다. 깊이 감지 하드웨어는 종종 값 비싸고, 예를 들어, 제한된 범위를 가지며 실내에서만 사용될 수 있는 것과 같이 그 사용이 제한적이다. 그러나 단일 RGB 이미지로부터의 입력 데이터를 사용하면 삼각 측량의 가능성이 배제되기 때문에 추가 하드웨어 없이 3D 위치를 유도하는 것은 어렵다.

[0004] 수학적으로, 단일 이미지의 2D 위치로부터 사람 관절의 깊이를 결정하는 것은 잘못된 것(ill-posed)이지만, 이 문제를 해결하기 위해 이전의 그리고 상황적 정보를 사용하는 방법이 제안되었다. 현대의 방법은 일반적으로 머신 러닝 기반 모델을 사용하며 크게 두 가지 범주로 나뉜다. 첫째, 리프팅 방법(lifting methods)은 이전에 검출된 사람의 2D 관절 위치(또는 골격)를 수신(accept)하고 일반적으로 완전히 연결된 신경망 모델을 사용하여 가장 가능성이 높은 포즈(pose)에 대응하는 3D 위치를 생성한다. 둘째, 컨볼루션 신경망(convolutional neural network; CNN) 기반 방법은 학습된 2D 컨볼루션을 사용하여 3D 위치 또는 깊이에 대한 회귀 모델에 제공되는, 사람을 포함하는 입력 이미지로부터 피처(features)를 추출한다.

[0005] 리프팅 접근 방식은, 관절의 상대적 깊이를 정확히 결정하는 데 중요할 수 있는 예를 들어, 그림자, 폐색(occlusions), 조명 및 볼륨(예컨대, 상이한 관절들의 겹보기 크기(apparent scale))과 같은 시각적 단서(visual cues)가 2D 골격 데이터에 누락되어 있다는 사실로 인해 제한된다. CNN 기반 접근 방식은 이러한 단서에 액세스할 수 있지만, 추출된 이미지 피처는 (2D 컨볼루션이 2D 맵에서 동작하기 때문에) 출력의 3D 특성과 대조적으로 본질적으로 2D이다. 또한, CNN 접근 방식은 일반적으로 2D 관절 위치에 대한 선형적 지식을 갖지 않으며 그 대신 연관된 깊이를 추론하면서 이러한 위치를 추론해야 한다.

[0006] 부분적으로 이러한 제한으로 인해, 기존 접근 방식은 특히 어려운 포즈(poses), 환경 및 조건에서 제한된 정확도와 견고성을 종종 보인다. 따라서 사람 관절 깊이의 단안 RGB 추정에 대한 더 나은 접근 방식이 필요하다.

발명의 내용

[0007] 물체의 이미지 세그먼트 및 이미지 세그먼트 상의 골격 관절의 2D 위치를 포함하는 2D 관절 히트맵(joint heatmaps)으로부터 골격 관절의 3D 관절 위치를 추정하기 위한 시스템 및 방법이 제공된다. 이것은 이미지 세그먼트 및 2D 관절 히트맵을, 적어도 하나의 3D 컨볼루션 계층 블록을 포함하는 컨볼루션 신경망에 적용하는 것을 포함하며, 여기서 2D 해상도는 각 3D 컨볼루션 계층에서 감소되고 깊이 해상도는 확장되어 각 관절에 대한 추정된 깊이를 생성한다. 각 관절 종류의 2D 위치를 관절 종류의 추정된 깊이와 결합하는 것은 골격 관절의 추정된

3D 관절 위치를 생성한다.

도면의 간단한 설명

본 개시의 바람직한 실시예만을 예시로서 예시하는 도면에서,

도 1은 3D 포즈 추정 시스템의 실시예의 고수준 아키텍처의 표현이다.

도 2는 3D CNN 블록을 통한 공간 차원 크기 감소 및 깊이 차원 크기 증가의 예의 표현이다.

도 3은 1개의 2D CNN 블록 및 4개의 3D CNN 블록을 도시하는 깊이 히트맵 추정기의 실시예의 개략적 표현이다.

도 4는 4개의 3D CNN 블록과 0개의 2D CNN 블록을 도시하는 깊이 히트맵 추정기의 실시예의 개략적 표현이다.

도 5는 2개의 잔차 유닛의 2D CNN 블록 및 5개의 3D CNN 블록을 갖는 깊이 히트맵 추정기의 실시예의 개략적 표현이다.

도 6은 각각 하나의 컨볼루션 계층을 포함하는 하나의 2D CNN 블록과 3개의 3D CNN 블록을 갖는 깊이 히트맵 추정기의 실시예의 개략적 표현이다.

도 7은 예시 입력 이미지, 대응 입력 히트맵의 컬러화된 표현, 및 골격으로서 묘사된 대응 3D 관절 위치의 예시의 표현이다.

도 8은 측위(localize)될 복수의 관절 종류에 대응하는 복수의 출력 깊이 히트맵(깊이 범위는 수평임)의 예이다.

발명을 실시하기 위한 구체적인 내용

도 1을 참조하면, 사람 관절 3D 추정 시스템(110)은, 디지털 이미지를 캡처하기 위한 캡처 디바이스(120), 이미지에서 사람을 검출하고 잘라내기(crop)하며 이미지에서 사람 관절 위치를 측위할 수 있는 2D 골격 검출기(130), 측위될 각 관절 종류에 대해 위치 히트맵을 생성할 수 있는 2D 관절 히트맵 생성기(140), 측위된 관절의 각 종류에 대해 1D 깊이 히트맵을 생성할 수 있는 깊이 히트맵 추정기(150), 및 깊이 히트맵으로부터 관절의 각 종류에 대해 깊이 값을 선택하고 이 정보를 2D 관절 위치와 결합하여 3D 관절 위치를 생성할 수 있는 3D 관절 구성기(160)를 포함할 수 있다.

캡처 디바이스(120)는 예를 들어, 디지털 카메라, 웹캠, 모바일 디바이스, 비디오 카메라, 스트리밍 소스(예컨대, 웹으로부터 스트리밍되는 비디오 또는 이미지) 또는 예를 들어, 이미지 또는 비디오 데이터베이스와 같은 저장된 이미지 콘텐츠와 같은, 디지털 이미지 또는 비디오 프레임을 획득하기 위한 임의의 디바이스 또는 방법을 표현할 수 있다. 이미지 형식은 3채널 RGB이거나 일부 다른 컬러 공간(예컨대, sRGB, CMYK)으로 인코딩될 수 있다. 이미지는 예를 들어, 깊이 센서 또는 적외선 센서로부터와 같은 추가 이미지 채널을 포함하거나 포함하지 않을 수 있다.

2D 골격 검출기(130)는 이미지에서 한 명 이상의 사람을 검출할 수 있다. 각각의 검출된 사람에 대해, 2D 골격 검출기는 그 사람을 중심으로 잘라내어진 이미지를 출력할 수 있다. 각 출력된 잘라내어진 이미지는 특정 해상도(예컨대, 256x256 픽셀)로 크기가 조정될 수 있으며, 이는 검출된 사람의 입력 해상도 또는 크기와 무관할 수 있다.

각각의 검출된 사람에 대해, 2D 골격 검출기는 미리 정의된 복수 종류의 사람 관절의 검출된 위치, 예를 들어, 2D 좌표를 출력할 수도 있다. 관절은 해부학적 랜드마크일 수 있다. 예를 들어, 2D 관절 검출기는 주요 관절의 모든 가시적 인스턴스(visible instances)의 위치를 측위할 수 있다. 여기에는 왼쪽 손목, 오른쪽 손목, 왼쪽 무릎, 목 기저부를 포함하여 22가지 종류의 관절이 포함될 수 있다. 2D 관절 검출기는 이러한 검출된 관절을, 검출된 사람에 의해 그룹화된 복수의 좌표로서 제공할 수 있다. 각 검출된 사람에 대한 관절의 좌표는 하나의 숫자 배열로 제공될 수 있다.

2D 골격 검출기는 예를 들어, OpenPose 또는 wrnchAI로부터의 기존 소프트웨어 모듈, 또는 사람을 검출하고 입력 이미지에서 사람 관절을 측위할 수 있는 일부 다른 소프트웨어 또는 하드웨어 모듈로 구현될 수 있다.

2D 관절 히트맵 생성기(140)는 한 사람에 대한 복수의 검출된 관절 위치를 수신하고 각 관절 종류에 대해 히트맵 이미지를 생성할 수 있다. 한 종류의 관절에 대한 히트맵 이미지는 그러한 종류의 복수의 검출된 관절의 위치를 해당 2D 관절 위치에서 측위된 이미지 내의 대응하는 복수의 마커(markers)로서 전달할 수 있다. 마커는

가우스 커널일 수 있다. 예를 들어, 위치는 8개의 픽셀 또는 일부 다른 크기의 표준 편차를 포함할 수 있다. 위치는 플랫 디스크(flat disk) 또는 임의의 다른 측위된 커널(kernel) 이미지의 형태일 수 있다.

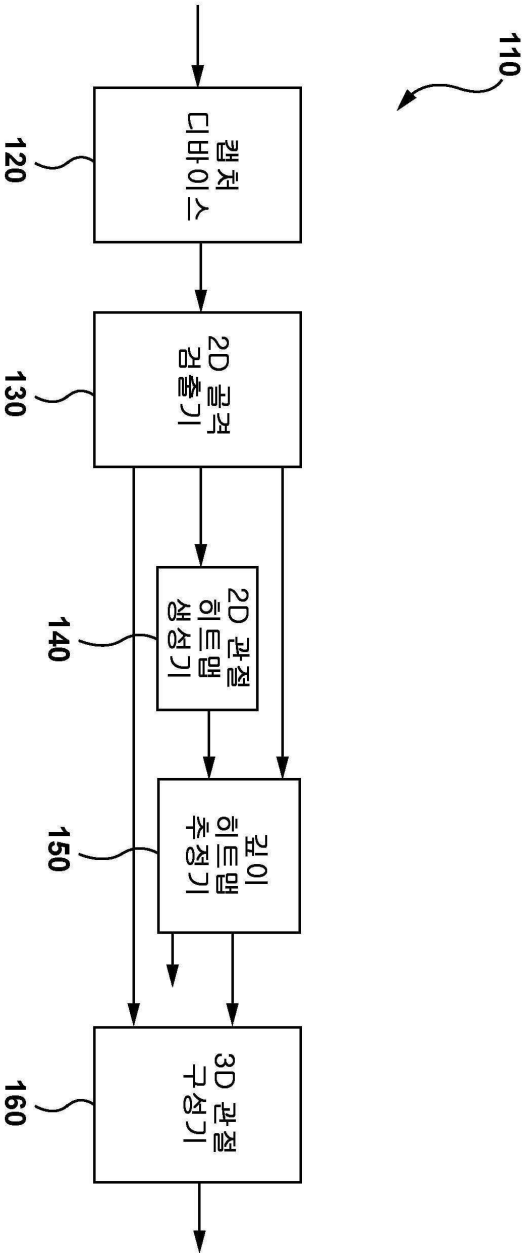
- [0015] 그 대신에 2D 관절 히트맵 생성기(140)는 해당 모듈의 사용 가능한 출력에 따라 2D 골격 검출기(130)에 의해 직접 생성된 2D 관절 히트맵을 출력할 수 있다. 예를 들어, 일부 2D 골격 검출기는 2D 관절 히트맵을 직접 출력할 수 있다. 이 경우 2D 관절 히트맵 생성기는 2D 골격 검출기의 부품 또는 하위 모듈일 수 있다.
- [0016] 2D 관절 히트맵 생성기에 의해 출력되는 각각의 히트맵 이미지는 관절의 한 종류에 대응하여 단일 채널에 인코딩될 수 있다. 해당 채널의 각 픽셀 값은 해당 픽셀에서 주어진 종류의 관절이 검출될 가능성을 근사할 수 있다. 히트맵 이미지의 해상도는 예를 들어, 256x256개의 픽셀과 같이 2D 골격 검출기에 의해 검출된 한 사람의 잘라내기된 이미지와 동일한 해상도일 수 있다.
- [0017] 깊이 히트맵 추정기(150)는 한 사람의 유사성을 중심으로 한 이미지를 그 사람의 미리 정의된 복수 종류의 관절의 위치를 나타내는 2D 히트맵 이미지와 함께 입력으로서 수신하여 복수의 1D 깊이 히트맵을 출력할 수 있다. 깊이 히트맵의 개수는 입력된 2D 히트맵의 개수에 대응할 수 있다. 이는 2D 골격 검출기에 의해 검출된 관절의 종류의 수와도 대응할 수 있다. 각 깊이 히트맵은 특정 고정점을 기준으로 한 관절 종류의 깊이 값 범위를 나타낼 수 있다. 이 고정점은 사람의 엉덩이의 중심, 카메라로부터 가장 먼 관절, 또는 일부 다른 기준점이 될 수 있다. 도 8을 참조하면, 깊이 히트맵은 픽셀의 1D 어레이(예컨대, 256개의 픽셀)로 저장될 수 있으며, 각 픽셀은 상대적 깊이 값에 대응한다. 각 깊이 히트맵의 각 픽셀 값은 대응 관절이 해당 상대 깊이에 있을 가능성을 나타낼 수 있다.
- [0018] 깊이 히트맵 추정기는 트레이닝된 머신 러닝 모델로 구현될 수 있다. 이 모델은 컨볼루션 신경망(convolutional neural network; CNN), 또는 대안적으로 RNN, 랜덤 포레스트(random forest), 또는 심층 신경망(deep neural network)일 수 있다. CNN 모델의 아키텍처는 입력 데이터로부터 3D 피쳐(예컨대, 객체 경계, 패색, 공간적 순서, 텍스처 및 컬러 패턴, 조명 피쳐)를 추출할 수 있는 3D 컨볼루션 계층을 포함할 수 있다. CNN 모델은 2D 컨볼루션 계층, 잔차 단위, 제한된 선형 활성화 단위(ReLU 또는 예를 들어, 시그모이드 단위와 같은 다른 활성화), 일괄 정규화 계층(batch normalization layers), 최대 풀링 계층(max pooling layers), 및 재성형 계층(reshape layers)도 포함할 수 있다. 이러한 계층은 예를 들어, 2D CNN 블록 및 3D CNN 블록과 같은 논리적 블록으로 조직화될 수 있으며 종단 간 사슬 연결된다(chained).
- [0019] CNN은, 각 블록에서 2D 해상도가 감소되고 그에 따라 깊이 해상도가 확장되도록 피라미드 방식으로 조직화될 수 있으며, 메모리의 효율적이고 제한된(bounded) 사용을 유지하면서 증가하는 충실도로 3D 피쳐를 추출한다.
- [0020] 2D CNN 블록은 낮은 수준의 이미지 피쳐를 추출하고 공간 해상도의 초기 감소(즉, 풀링)를 수행하여 수용 필드(receptive field)를 확장할 수 있다. 2D CNN 블록은 2D 컨볼루션 계층, 잔차 유닛, 최대 풀링 계층(예컨대, 2x2x1 커널 크기를 가짐) 및 최종 2D 컨볼루션 계층을 포함할 수 있다. 일괄 정규화 계층은 일괄 정규화를 부과하기 위해 각 컨볼루션 계층 뒤에 삽입될 수 있다. 컨볼루션 계층은 예를 들어, 7x7 커널 크기와 보폭(stride) 2를 가질 수 있다. 잔차 유닛은 스킵 연결을 갖는 3개의 컨볼루션 계층을 가질 수 있다.
- [0021] 2D CNN 블록 포함의 중요성은, 이 블록에 의해 수행되는 공간적 해상도의 감소가 후속 계층, 예를 들어, 상대적으로 계산 비용이 많이 들 수 있는 3D 컨볼루션 계층의 메모리 및 계산 비용을 감소시킬 수 있기 때문에 메모리 및 계산 고려 사항에 따라 달라질 수 있다.
- [0022] 3D CNN 블록은 3D 컨볼루션, 및 ReLU, 뒤이어 또 다른 3D 컨볼루션 및 ReLU, 뒤이어 최대 풀링 계층 및 재성형 계층을 포함할 수 있다. 일괄 정규화 계층은 일괄 정규화를 부과하기 위해 각 컨볼루션 계층 뒤에 삽입될 수 있다. 각 3D 컨볼루션은 계산 비용과 각 후속 계층에 필요한 학습 가능한 파라미터의 수를 감소시키기 위해 3x3x3의 커널 크기를 가질 수 있다.
- [0023] 3D CNN 블록의 각 최대 풀링 계층은 입력된 공간 차원 H 및 W의 크기를 예를 들어, 각각 2배(a factor of 2)만큼 감소시킬 수 있지만, 깊이 차원 D를 유지할 수 있다. 각 재성형 계층은 풀링 계층에 의해 생성된 채널의 수를 감소시키고 그에 따라 깊이 차원의 크기를 예를 들어, 4배만큼 증가시킬 수 있다. 따라서, 도 2를 참조하면, 각각의 3D CNN 블록은 D 차원을 비례적으로 확장하면서 H 차원 및 W 차원을 효과적으로 감소시킬 수 있다.
- [0024] 도 3을 참조하면, 깊이 히트맵 추정기는 2D CNN 블록, 하나 이상의 3D CNN 블록, 예를 들어, 4개의 블록, 이어서 하나(또는 그 이상)의 최종 3D 컨볼루션 계층 및 최종 최대 풀링 계층을 포함할 수 있다. 최종 3D 컨볼루션 계층은 추정된 관절의 종류의 수와 동일한 수의, 예를 들어, 22개의 채널을 출력할 수 있다. 최종 최대 풀링 계

층은 원래의 공간 차원 W 및 H 를 각각 1로 감소시키고 각 관절 채널에 대해 확장된 깊이 차원만 남길 수 있다.

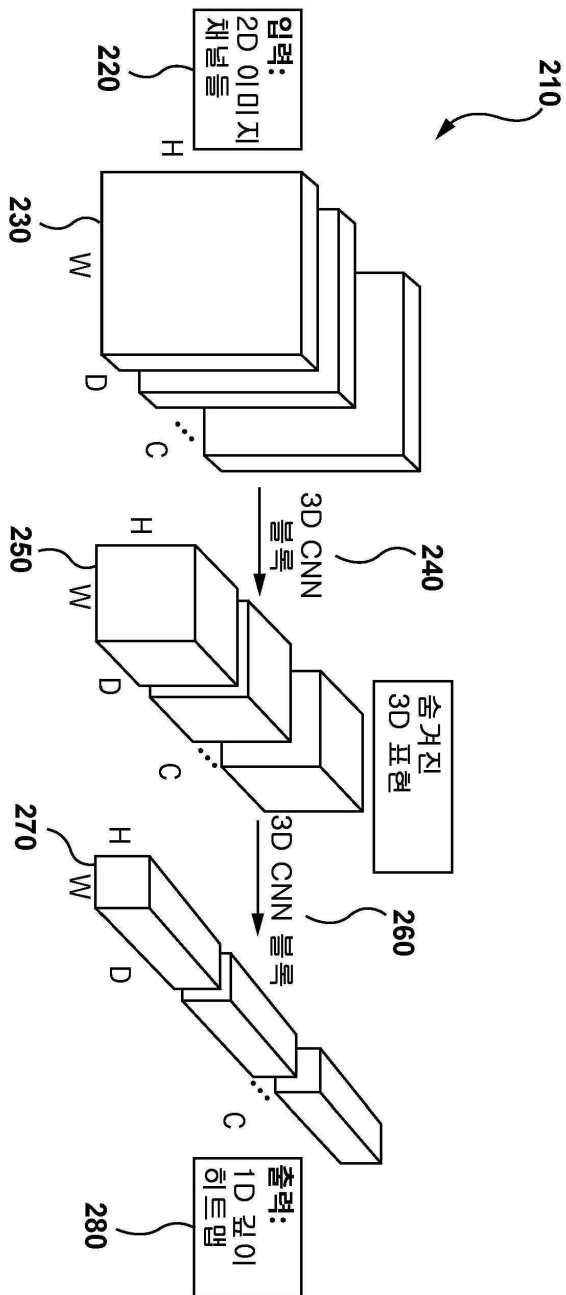
- [0025] 깊이 히트맵 추정기의 아키텍처는 메모리 제약, 런타임 제약, 트레이닝 데이터베이스 크기, 및 다른 유사한 요인을 포함한, 작업의 특정 요건에 따라 달라질 수 있다. 예를 들어, 도 4를 참조하면, 깊이 히트맵 추정기는 2D 블록을 포함하지 않을 수 있다. 도 5를 참조하면, 2D CNN 블록은 하나보다 많은 잔차 단위를 포함할 수 있고 깊이 히트맵 추정기는 4개보다 많은 3D 블록을 포함할 수 있다. 도 6을 참조하면, 깊이 히트맵 추정기는 4개 미만의 3D 블록을 포함할 수 있고 각 3D 블록은 상이한 수들의 적층된 3D 컨볼루션 계층을 포함할 수 있다.
- [0026] 각 블록에 더 많은 3D 컨볼루션 계층이 있으면 더 복잡한 3D 피처를 추출할 수 있지만 트레이닝 중에 더 많은 계산이 필요하고 메모리 풋프린트(memory footprint)가 커진다. 따라서 주어진 실시예에서 블록당 컨볼루션 계층의 특정 수는 애플리케이션의 요건에 따라 달라질 수 있다.
- [0027] 참조의 편의를 위해, 도 3, 도 4, 도 5 및 도 6에는 ReLU 및 일괄 정규화 계층이 존재할 수 있지만 도시되지는 않는다.
- [0028] 깊이 히트맵 추정기의 학습 가능한 파라미터는 적절한 트레이닝 세트로 트레이닝될 수 있다. 트레이닝은 역전파(back propagation) 또는 또 다른 적절한 학습 알고리즘을 사용될 수 있다. 예를 들어, RMSProp, 기울기 하강법, ADAM 등의 솔버 방법(solver method)을 사용할 수 있다. 예를 들어, TensorFlow, Caffe 또는 PyTorch 또는 유사한 프레임워크와 같은 트레이닝 프레임워크가 사용될 수 있다. 트레이닝 세트는 각각 라벨 표기된 3D 관절 위치와 연관된 상이한 포즈들, 설정들 및 조명 조건들에 있는 예를 들어, 수천 개의 사람 이미지의 세트를 포함할 수 있다. 예를 들어, COCO 데이터 세트가 사용될 수 있다.
- [0029] 3D 관절 구성기(160)는 예를 들어, 2D 골격 검출기(130)로부터의 한 명의 검출된 사람에 대한 2D 관절 위치, 및 깊이 히트맵 추정기(150)에 의해 생성된 대응 관절 깊이 히트맵을 수신하고, 3D 관절 위치를 출력할 수 있다. 3D 관절 위치의 깊이 성분은 argmax 함수를 통해 깊이 히트맵으로부터 계산될 수 있다. 대안적으로, 가중 합 함수 또는 피크 검출 기법이 사용될 수 있다. 입력된 2D 관절 위치와 계산된 깊이 값을 결합하여 3D 관절 위치를 생성할 수 있다.
- [0030] 사람 관절 3D 추정 시스템의 일 실시예로부터 생성된 복수의 3D 관절 위치의 예가 도 7에 묘사되어 있다. 인체(710)를 나타내는 입력 이미지는 예를 들어, 2D 관절 히트맵 생성기(140)에 의해 생성된 것과 같은 2D 관절 히트맵(720)과 함께 제공된다. 히트맵(720)에 표현되지는 않았지만 예를 들어, 왼쪽 손목과 같은 관절의 각 종류는 별도의 히트맵을 가질 것이다. 모든 검출된 관절에 대한 히트맵은 오른쪽 관절이 빨간색으로 컬러화되고, 왼쪽 관절이 파란색으로 컬러화된, 컬러화된 표현(720)에 오버레이된다.
- [0031] 깊이 히트맵 추정기에 의해 생성된 깊이 히트맵의 예는 도 8에 묘사된다. 깊이 추정기는 각 관절 유형에 대해 별도의 깊이 히트맵을 생성한다. 각 관절의 깊이 범위는 수평축에 표시된다. 주어진 히트맵을 따라 각 지점의 값(즉, 어두움)은 대응 관절이 해당 깊이에 있다는 신뢰도를 나타낼 수 있다. 모든 깊이의 신뢰도 값은 주어진 관절에 대해 0일 수 있으며, 이는 가시적이지 않은 관절에 대응할 수 있다. 추정된 깊이 값은 상이한 깊이들에서의 신뢰도 값들을 합산(aggregate)하여 유도될 수 있다.
- [0032] 2D 관절 히트맵과 각 관절 종류에 대한 깊이 추정을 결합하여 각 관절에 대한 3D 위치가 추정될 수 있다. 이들 3D 위치에 기초한 골격의 렌더링(740)이 도시되어 있다.
- [0033] 이러한 설명된 시스템 모듈은 별도의 소프트웨어 모듈, 별도의 하드웨어 유닛 또는 하나 이상의 소프트웨어 또는 하드웨어 컴포넌트의 부분일 수 있다. 예를 들어, 소프트웨어 모듈은 예를 들어, Caffe, TensorFlow 또는 Torch로부터의 소프트웨어를 사용하여 생성되고, 예를 들어, CPU, GPU와 같은 컴퓨터 하드웨어에서 실행되거나 FPGA에서 구현되는, 적절한 모듈과 함께 예를 들어, Python 프로그래밍 언어, C++ 또는 C#과 같은 컴퓨터 언어로 작성된 명령어로 이루어질 수 있다. 시스템은 데스크톱, 휴대폰 또는 예를 들어, 소프트웨어 및 활동 데이터를 보유하기 위한 적절한 메모리를 포함하는 임베디드 시스템의 일부와 같은 다른 플랫폼에서 실행될 수 있다. 시스템은 캡처 디바이스와 통합되거나 접속될 수 있다.
- [0034] 이와 같이 본 개시의 다양한 실시예가 예로서 상세하게 설명되었지만, 본 개시의 범위를 벗어나지 않고 변형 및 수정이 이루어질 수 있음은 당업자에게 명백할 것이다. 본 개시는 첨부된 청구항들의 범위 내에 속하는 모든 그러한 변형 및 수정을 포함한다.

도면

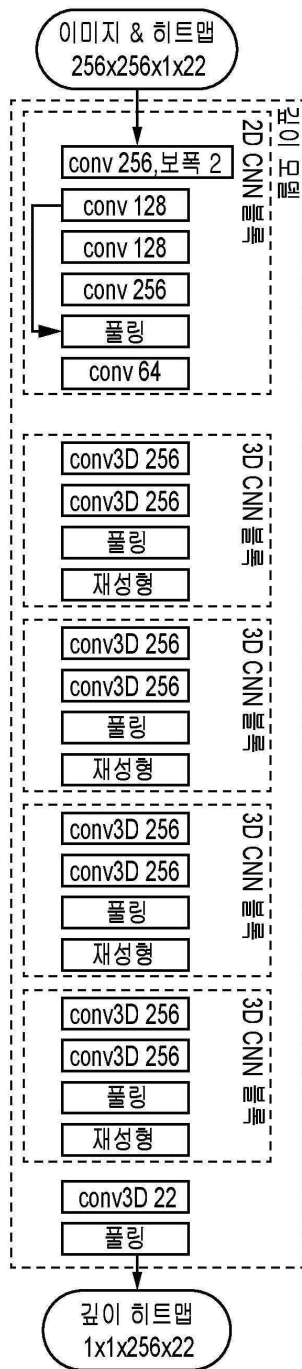
도면1



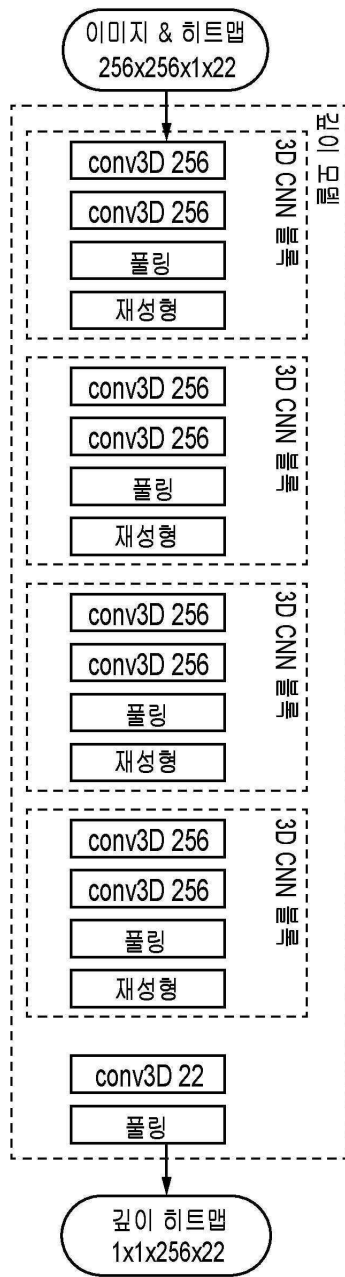
도면2



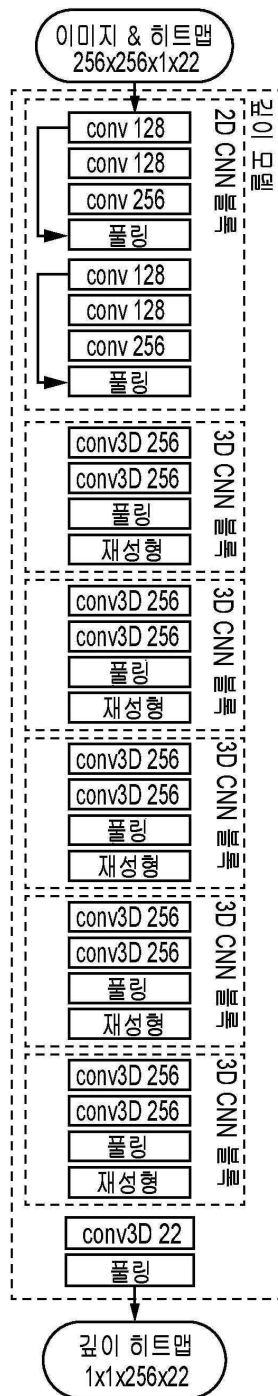
도면3



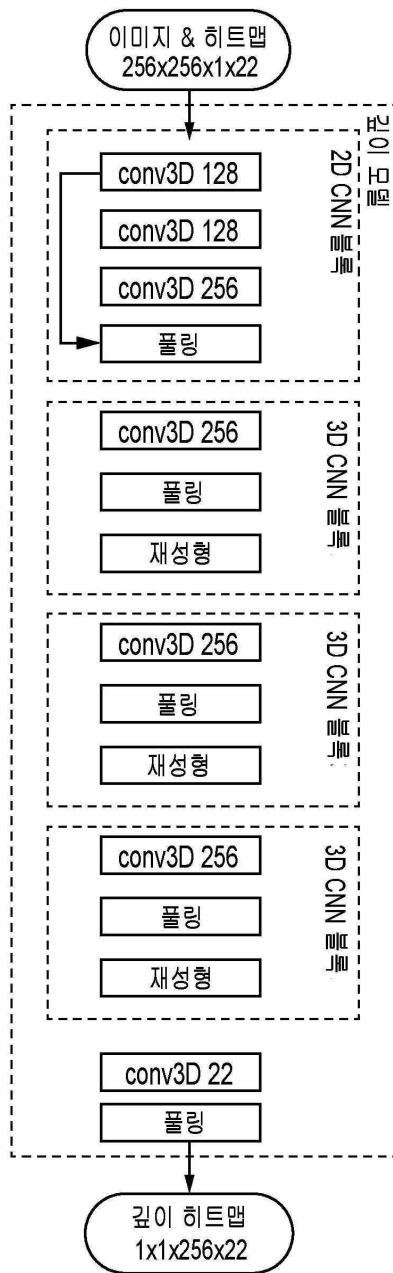
도면4



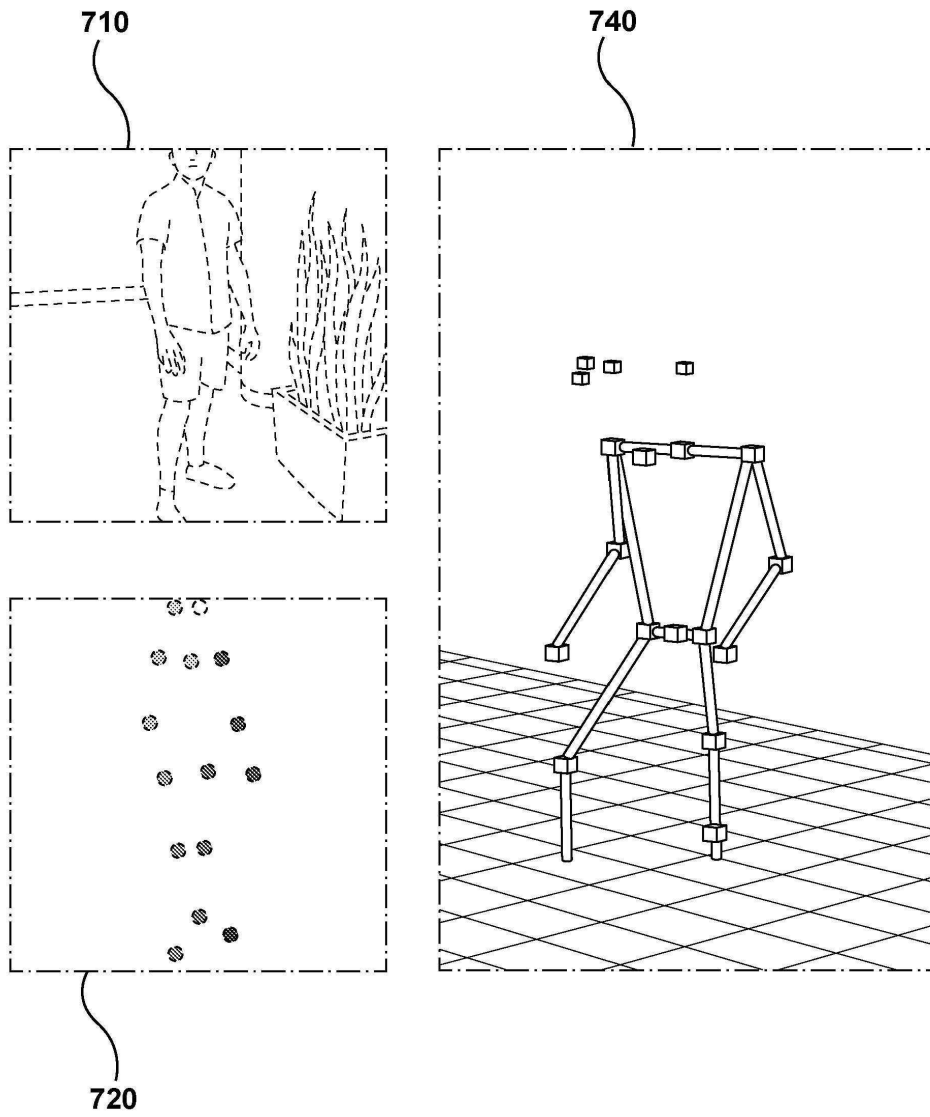
도면5



도면6



도면7



도면8

