(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0083008
(43) 공개일자 2024년06월11일

(51) 국제특허분류(Int. Cl.)
G06T 17/20 (2006.01) *G06N 3/0464* (2023.01)
G06N 3/048 (2023.01) *G06N 3/08* (2023.01)
G06T 15/08 (2011.01) *G06T 7/593* (2017.01)
(52) CPC특허분류
G06T 17/20 (2013.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0131489
(22) 출원일자 2023년10월04일
심사청구일자 없음
(30) 우선권주장
1020220166858 2022년12월02일 대한민국(KR)

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
이희경
대전광역시 유성구 가정로 218
엄기문
대전광역시 유성구 가정로 218
(뒷면에 계속)
(74) 대리인
최윤서

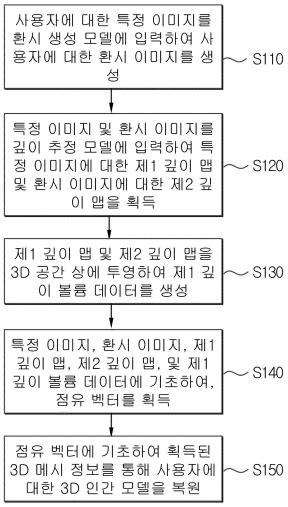
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **깊이 투영 기반의 3D 모델 복원 신경망을 이용하는 방법 장치**

(57) 요 약

3D 인간 모델을 복원하는 방법 및 장치가 개시된다. 본 개시의 일 실시예에 따른, 장치에 의해 수행되는, 3D 인간 모델을 복원하는 방법은, 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하는 단계; 상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하는 단계; 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하는 단계; 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하는 단계; 및 상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 사용자에 대한 3D 인간 모델을 복원하는 단계를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/048 (2023.01)
G06N 3/084 (2023.01)
G06T 15/08 (2013.01)
G06T 7/593 (2017.01)
G06T 2207/20084 (2013.01)
G06T 2207/30196 (2013.01)
(72) 발명자
송대영
대전광역시 유성구 대학로 99, 충남대학교

조동현
대전광역시 유성구 대학로 99, 충남대학교

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152526
과제번호	2018-0-00207
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	ETRI연구개발지원사업
연구과제명	이머시브 미디어 전문연구실
기 여 율	1/1
과제수행기관명	ETRI
연구기간	2022.01.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

장치에 의해 수행되는, 3D 인간 모델을 복원하는 방법에 있어서, 상기 방법은:

사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하는 단계;

상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하는 단계;

상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하는 단계;

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하는 단계; 및

상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 사용자에 대한 3D 인간 모델을 복원하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 특정 이미지는, 상기 사용자의 앞모습을 나타내는 이미지이고,

상기 환시 이미지는, 상기 사용자의 뒷모습을 나타내는 이미지인, 방법.

청구항 3

제1항에 있어서,

상기 환시 생성 모델은, 다운 샘플링(down-sampling) 합성곱 레이어, 업 샘플링(up-sampling) 합성곱 레이어, 잔차(residual) 합성곱 레이어, 및 활성화 함수 중의 적어도 하나를 포함하는, 방법.

청구항 4

제1항에 있어서,

상기 깊이 추정 모델은, 깊이 추정 인코딩부 및 깊이 추정 디코딩부를 포함하고,

상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하는 단계는,

상기 특정 이미지 및 상기 환시 이미지를 상기 깊이 추정 인코딩부에 입력하여 제1 인코딩 특징맵을 획득하는 단계; 및

상기 제1 인코딩 특징맵 및 상기 사용자의 키 대비 깊이 비율을 상기 깊이 추정 디코딩부에 입력하여 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하는 단계를 포함하는, 방법.

청구항 5

제1항에 있어서,

상기 점유 벡터를 획득하는 단계는,

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 2D 기하 인코딩부에 입력하여 제2 인코딩 특징맵을 획득하는 단계;

상기 제1 깊이 볼륨 데이터를 3D 기하 인코딩부에 입력하여 제3 인코딩 특징맵을 획득하는 단계; 및

상기 제2 인코딩 특징 맵 및 상기 제3 인코딩 특징 맵을 점유 추정 모델에 입력하여 상기 점유 벡터를 획득하는 단계를 포함하는, 방법.

청구항 6

제1항에 있어서,

상기 점유 벡터에 대해 마칭 큐브(marching cube) 알고리즘을 적용하여 상기 3D 메시 정보가 획득되는, 방법.

청구항 7

제1항에 있어서,

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 제2 깊이 볼륨 데이터에 기초하여, 색상 예측 정보가 획득되고,

상기 색상 예측 정보 및 상기 3D 메시 정보에 기초하여 상기 3D 인간 모델이 복원되는, 방법.

청구항 8

제5항에 있어서,

상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수의 값이 산출되고,

상기 산출된 손실 함수의 값이 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각에 대해 역전파(back-propagation)됨에 따라 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각이 학습되는, 방법.

청구항 9

제8항에 있어서,

상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수에 입력될 변수 데이터는 데이터 셋 내에 수집되는, 방법.

청구항 10

3D 인간 모델을 복원하는 장에 있어서, 상기 장치는:

하나 이상의 메모리; 및

상기 하나 이상의 프로세서를 포함하고,

상기 하나 이상의 프로세서는:

사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하고;

상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하고;

상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하고;

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하고; 및

상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 사용자에 대한 3D 인간 모델을 복원하도록 설정되는, 장치.

청구항 11

제10항에 있어서,

상기 특정 이미지는, 상기 사용자의 앞모습을 나타내는 이미지이고,

상기 환시 이미지는, 상기 사용자의 뒷모습을 나타내는 이미지인, 장치.

청구항 12

제10항에 있어서,

상기 환시 생성 모델은, 다운 샘플링(down-sampling) 합성곱 레이어, 업 샘플링(up-sampling) 합성곱 레이어, 잔차(residual) 합성곱 레이어, 및 활성화 함수 중의 적어도 하나를 포함하는, 장치.

청구항 13

제10항에 있어서,

상기 깊이 추정 모델은, 깊이 추정 인코딩부 및 깊이 추정 디코딩부를 포함하고,

상기 하나 이상의 프로세서는,

상기 특정 이미지 및 상기 환시 이미지를 상기 깊이 추정 인코딩부에 입력하여 제1 인코딩 특징맵을 획득하고; 및

상기 제1 인코딩 특징맵 및 상기 사용자의 키 대비 깊이 비율을 상기 깊이 추정 디코딩부에 입력하여 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하도록 설정되는, 장치.

청구항 14

제10항에 있어서,

상기 하나 이상의 프로세서는,

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 2D 기하 인코딩부에 입력하여 제2 인코딩 특징맵을 획득하고;

상기 제1 깊이 볼륨 데이터를 3D 기하 인코딩부에 입력하여 제3 인코딩 특징맵을 획득하고; 및

상기 제2 인코딩 특징 맵 및 상기 제3 인코딩 특징 맵을 점유 추정 모델에 입력하여 상기 점유 벡터를 획득하도록 설정되는, 장치.

청구항 15

제10항에 있어서,

상기 점유 벡터에 대해 마칭 큐브(marching cube) 알고리즘을 적용하여 상기 3D 메시 정보가 획득되는, 장치.

청구항 16

제10항에 있어서,

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 제2 깊이 볼륨 데이터에 기초하여, 색상 예측 정보가 획득되고,

상기 색상 예측 정보 및 상기 3D 메시 정보에 기초하여 상기 3D 인간 모델이 복원되는, 장치.

청구항 17

제14항에 있어서,

상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수의 값이 산출되고,

상기 산출된 손실 함수의 값이 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각에 대해 역전파(back-propagation)됨에 따라 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각이 학습되는, 장치.

청구항 18

제17항에 있어서,

상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수에 입력될 변수 데이터는 데이터 셋 내에 수집되는, 장치.

청구항 19

하나 이상의 명령을 저장하는 하나 이상의 비-일시적(non-transitory) 컴퓨터 판독가능 매체로서,

상기 하나 이상의 명령은 하나 이상의 프로세서에 의해서 실행되어, 3D 인간 모델을 복원하는 장치가:

사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하고;

상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하고;

상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하고;

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하고; 및

상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 3D 인간 모델을 복원하도록 제어되는, 컴퓨터 판독가능 매체.

청구항 20

3D 인간 모델을 복원하는 시스템에 있어서, 상기 시스템은:

사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 환시 생성부;

상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하는 깊이 추정부;

상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하는 깊이 투영부;

상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하는 모델 점유 추정부; 및

상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 3D 인간 모델을 복원하는 복원부를 포함하는, 시스템.

발명의 설명

기 술 분 야

[0001] 본 개시는 3D 모델 복원 기술에 관한 것이며, 더욱 상세하게는 입력 사람 이미지에 기초하여 3D 모델을 복원하는 방법을 수행하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 종래에는 사람에 대한 3D 모델을 생성하기 위하여, 3D 스캐닝(Scanning) 장비 등 복잡한 장비를 360도 방향으로 세팅해야 했다. 사람에 대한 3D 모델이 엔터테인먼트(Entertainment), 화상 회의(Video Conference) 등 다양한 분야에 적용된다는 것을 고려할 때, 종래의 기술은 이동성이 떨어질 뿐만 아니라 매번 복잡한 장비의 재설치를 요구한다는 점에서 경제성이 떨어진다는 문제점이 존재하였다.

[0003] 이를 해결하기 위하여, 단일 사람 이미지에 대해 딥러닝 신경망 방법을 적용하여 3D 모델을 복원하는 시도들이 존재하였다.

[0004] PIFu 등의 완전 연결 계층(Fully-connected Layer)으로 구성된 딥러닝 기반의 모델들은 3D 모델의 앞모습은 비교적 자연스럽게 복원하였으나, 3D 모델의 반을 구성하는 뒷모습(unseened surface)에 대해서는 원활하게 복원하지 못한다는 문제점이 존재하였다.

[0005] 종래의 딥러닝 기반 모델들은 오버스무딩(over-smoothing)을 유발하며 옷 결 등을 표현하지 못하였다. 이는 딥러닝 기반 모델의 학습 과정에서 뒷모습에 대한 정보 없이 손실 함수를 최적(최소)화하기 위해 스무딩을 수행함에 따라 관련 평균값이 작아지기 때문이다.

[0006] 뿐만 아니라, 완전 연결 계층 기반 신경망은 합성곱(Convolution) 신경망과는 다르게 이동 등가성(Translation Equivariance) 및 지역성(Locality)로 대표되는 귀납 편향(Inductive Bias)이라는 강력한 가정을 사용할 수 없다. 이런 상황에서, 적절한 조건(Conditioning) 없이 완전 연결 계층 기반 복원을 수행할 경우, 3D 모델을 상세하게 표현하기 어렵다는 문제점이 존재한다.

발명의 내용

해결하려는 과제

[0007] 본 개시는 상술된 문제점을 해결하기 위해 안출된 것으로서, 본 개시의 목적은 인간 이미지에 대해 딥러닝 알고리즘을 적용함으로써 인간 3D 모델을 효율적으로 생성하는 방법 및 장치를 제공함에 있다.

[0008] 또한, 본 개시의 목적은 인간 3D 모델의 뒷면 및 옆면을 귀납 편향 가정이 가능한 신경망 모델을 통해 효율적으로 복원하는 방법 및 장치를 제공함에 있다.

[0009] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0010] 본 개시의 일 실시예에 따른, 장치에 의해 수행되는, 3D 인간 모델을 복원하는 방법은, 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하는 단계; 상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하는 단계; 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하는 단계; 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하는 단계; 및 상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 사용자에 대한 3D 인간 모델을 복원하는 단계를 포함할 수 있다.

[0011] 그리고, 상기 특정 이미지는, 상기 사용자의 앞모습을 나타내는 이미지이고, 상기 환시 이미지는, 상기 사용자의 뒷모습을 나타내는 이미지일 수 있다.

[0012] 그리고, 상기 환시 생성 모델은, 다운 샘플링(down-sampling) 합성곱 레이어, 업 샘플링(up-sampling) 합성곱 레이어, 잔차(residual) 합성곱 레이어, 및 활성화 함수 중의 적어도 하나를 포함할 수 있다.

[0013] 그리고, 상기 깊이 추정 모델은, 깊이 추정 인코딩부 및 깊이 추정 디코딩부를 포함하고, 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하는 단계는,

[0014] 상기 특정 이미지 및 상기 환시 이미지를 상기 깊이 추정 인코딩부에 입력하여 제1 인코딩 특징맵을 획득하는 단계; 및 상기 제1 인코딩 특징맵 및 상기 사용자의 키 대비 깊이 비율을 상기 깊이 추정 디코딩부에 입력하여 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하는 단계를 포함할 수 있다.

[0015] 그리고, 상기 점유 벡터를 획득하는 단계는, 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 2D 기하 인코딩부에 입력하여 제2 인코딩 특징맵을 획득하는 단계; 상기 제1 깊이 볼륨 데이터를 3D 기하 인코딩부에 입력하여 제3 인코딩 특징맵을 획득하는 단계; 및 상기 제2 인코딩 특징 맵 및 상기 제3 인코딩 특징 맵을 점유 추정 모델에 입력하여 상기 점유 벡터를 획득하는 단계를 포함할 수 있다.

[0016] 그리고, 상기 점유 벡터에 대해 마칭 큐브(marching cube) 알고리즘을 적용하여 상기 3D 메시 정보가 획득될 수 있다.

[0017] 그리고, 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 제2 깊이 볼륨 데이터에 기초하여, 색상 예측 정보가 획득되고, 상기 색상 예측 정보 및 상기 3D 메시 정보에 기초하여 상기 3D 인간 모델이 복원될 수 있다.

[0018] 그리고, 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수의 값이 산출되고, 상기 산출된 손실 함수의 값이 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각에 대해 역전파(back-propagation)됨에 따라 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각이 학습될 수 있다.

[0019] 그리고, 상기 환시 추정 모델, 상기 깊이 추정 모델 및 상기 점유 추정 모델 각각을 학습시키기 위한 손실 함수에 입력될 변수 데이터는 데이터 셋 내에 수집될 수 있다.

[0020] 본 개시의 일 실시예에 따른, 3D 인간 모델을 복원하는 장치는, 하나 이상의 메모리; 및 상기 하나 이상의 프로세서를 포함하고, 상기 하나 이상의 프로세서는: 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하고; 상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하고; 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하고; 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하고; 및 상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 사용자에 대한 3D 인간 모델을 복원하도록 설정될 수 있다.

[0021] 그리고, 상기 하나 이상의 프로세서는, 상기 특정 이미지 및 상기 환시 이미지를 상기 깊이 추정 인코딩부에 입력하여 제1 인코딩 특징맵을 획득하고; 및 상기 제1 인코딩 특징맵 및 상기 사용자의 키 대비 깊이 비율을 상기 깊이 추정 디코딩부에 입력하여 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 획득하도록 설정될 수 있다.

[0022] 그리고, 상기 하나 이상의 프로세서는, 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 2D 기하 인코딩부에 입력하여 제2 인코딩 특징맵을 획득하고; 상기 제1 깊이 볼륨 데이터를 3D 기하 인

코딩부에 입력하여 제3 인코딩 특징맵을 획득하고; 및 상기 제2 인코딩 특징 맵 및 상기 제3 인코딩 특징 맵을 점유 추정 모델에 입력하여 상기 점유 벡터를 획득하도록 설정될 수 있다.

[0023] 본 개시의 일 실시예에 따른, 하나 이상의 명령을 저장하는 하나 이상의 비-일시적(non-transitory) 컴퓨터 판독가능 매체로서, 상기 하나 이상의 명령은 하나 이상의 프로세서에 의해서 실행되어, 3D 인간 모델을 복원하는 장치가: 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 생성하고; 상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하고; 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하고; 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하고; 및 상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 3D 인간 모델을 복원하도록 제어될 수 있다.

[0024] 그리고, 본 개시의 일 실시예로, 3D 인간 모델을 복원하는 시스템은, 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 상기 사용자에 대한 환시 이미지를 환시 생성부; 상기 특정 이미지 및 상기 환시 이미지를 깊이 추정 모델에 입력하여 상기 특정 이미지에 대한 제1 깊이 맵 및 상기 환시 이미지에 대한 제2 깊이 맵을 획득하는 깊이 추정부; 상기 제1 깊이 맵 및 상기 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성하는 깊이 투영부; 상기 특정 이미지, 상기 환시 이미지, 상기 제1 깊이 맵, 상기 제2 깊이 맵, 및 상기 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득하는 모델 점유 추정부; 및 상기 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 상기 3D 인간 모델을 복원하는 복원부를 포함할 수 있다.

[0025] 본 개시에 대하여 위에서 간략하게 요약된 특징들은 후술하는 본 개시의 상세한 설명의 예시적인 양상일 뿐이며, 본 개시의 범위를 제한하는 것은 아니다.

발명의 효과

[0026] 본 개시의 다양한 실시예에 의해, 인간 이미지에 대해 딥러닝 알고리즘을 적용함으로써 인간 3D 모델을 효율적으로 생성하는 방법 및 장치가 제공될 수 있다.

[0027] 또한, 본 개시의 다양한 실시예에 의해, 인간 3D 모델의 뒷면 및 옆면을 귀납 편향 가정이 가능한 신경망 모델을 통해 효율적으로 복원하는 방법 및 장치가 제공될 수 있다.

[0028] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0029] 도 1은 본 개시의 일 실시예에 따른, 깊이 투영 기반의 3D 모델 복원 신경망을 이용하는 방법을 설명하기 위한 순서도이다.

도 2는 본 개시의 일 실시예에 따른, 딥러닝 기반의 인간 3D 모델 복원 신경망 파이프라인을 설명하기 위한 도면이다.

도 3은 본 개시의 일 실시예에 따른, 환시 생성기의 구성을 설명하기 위한 도면이다.

도 4는 본 개시의 일 실시예에 따른, 잔차 합성곱 모듈의 구성을 설명하기 위한 도면이다.

도 5는 본 개시의 일 실시예에 따른, 깊이 추정 네트워크 구조를 설명하기 위한 도면이다.

도 6은 본 개시의 일 실시예에 따른, 깊이 맵 투영 절차를 설명하기 위한 도면이다.

도 7은 본 개시의 실시예에 따른 장치를 예시하는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 개시는 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 개시를 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다. 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다. 후술하는 예시적 실시예들에 대한 상세한

설명은, 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 실시예를 실시할 수 있기에 충분하도록 상세히 설명된다. 다양한 실시예들은 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 개시의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 실시예의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 예시적 실시예들의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다.

[0031] 본 개시에서 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0032] 본 개시의 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있으나, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어"있다거나 "직접 접속되어"있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0033] 본 개시의 실시예에 나타나는 구성부들은 서로 다른 특징적인 기능들을 나타내기 위해 독립적으로 도시되는 것으로, 각 구성부들이 분리된 하드웨어나 하나의 소프트웨어 구성단위로 이루어짐을 의미하지 않는다. 즉, 각 구성부는 설명의 편의상 각각의 구성부로 나열하여 포함한 것으로 각 구성부 중 적어도 두개의 구성부가 합쳐져 하나의 구성부로 이루어지거나, 하나의 구성부가 복수 개의 구성부로 나뉘어져 기능을 수행할 수 있고 이러한 각 구성부의 통합된 실시예 및 분리된 실시예도 본 개시의 본질에서 벗어나지 않는 한 본 개시의 권리범위에 포함된다.

[0034] 본 개시에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 개시를 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 개시에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다. 즉, 본 개시에서 특정 구성을 "포함"한다고 기술하는 내용은 해당 구성 이외의 구성을 배제하는 것이 아니며, 추가적인 구성이 본 개시의 실시 또는 본 개시의 기술적 사상의 범위에 포함될 수 있음을 의미한다.

[0035] 본 개시의 일부의 구성 요소는 본 개시에서 본질적인 기능을 수행하는 필수적인 구성 요소는 아니고 단지 성능을 향상시키기 위한 선택적 구성 요소일 수 있다. 본 개시는 단지 성능 향상을 위해 사용되는 구성 요소를 제외한 본 개시의 본질을 구현하는데 필수적인 구성부만을 포함하여 구현될 수 있고, 단지 성능 향상을 위해 사용되는 선택적 구성 요소를 제외한 필수 구성 요소만을 포함한 구조도 본 개시의 권리범위에 포함된다.

[0036] 이하, 도면을 참조하여 본 개시의 실시 형태에 대하여 구체적으로 설명한다. 본 명세서의 실시예를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 명세서의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략하고, 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0038] 본 개시에서 제안하는 시스템 및/또는 방법/장치(이하 '시스템'으로 간단히 표현함)는 깊이 투영 기반의 3D 모델 복원 신경망 기술에 관한 것이다.

[0039] 상술한 바와 같이, 종래의 딥러닝 기반 기술들은 3D 결과물의 절반을 이루는 뒷면에 대한 세밀한 표현을 고려하지 않거나, 단순 파라미터(parametric) 모델을 사용했기 때문에 3D 결과물의 완성도가 높지 않다는 문제점이 존재한다.

[0040] 또한, 종래의 딥러닝 기반 기술들은 일반적으로 단일 입력 영상을 이용하였다. 이에 따라, 종래의 딥러닝 기반 기술을 이용하여 출력된 3D 결과물에는 깊이 모호성(depth ambiguity)이 발생하며, 3D 결과물의 옆 모습이 제대로 복원되지 않는다는 문제점이 존재하였다.

[0041] 이를 해결하기 위하여, 파라미터 모델을 사용하여 3D 모델을 복원하는 방식이 고안되었다. 다만, 파라미터 모델은 통계적 템플릿을 기반으로 하는 나체 모델(naked model)이므로 피부로부터 거리가 먼 머리카락이나 치마 등의 옷은 제대로 표현되지 않는 문제점이 존재하였다.

[0042] 상술된 문제점을 극복하기 위하여, 이하에서는 파라미터 모델이 표현하지 못하는 부분을 귀납 편향 가정이 가능한 두 신경망으로 복원하는 방법 및 장치를 설명한다.

[0044] 도 1은 본 개시의 일 실시예에 따른, 깊이 투영 기반의 3D 모델 복원 신경망을 이용하는 방법을 설명하기 위한 순서도이다.

[0045] 장치는 사용자에 대한 특정 이미지를 환시(hallucination) 생성 모델에 입력하여 사용자에 대한 환시 이미지를 생성할 수 있다(S110).

[0046] 여기서, 특정 이미지는 사용자의 앞모습을 나타내는 이미지이고, 환시 이미지는 사용자의 뒷모습을 나타내는 이미지를 나타낼 수 있다. 그리고, 사용자는 3D 모델의 복원의 대상이 되는 주체를 통칭할 수 있다.

[0047] 환시 생성 모델은 특정 이미지를 기준으로 폐색 영역인 사용자의 뒷모습을 나타내는 이미지를 생성하도록 학습된 모델을 의미한다. 일 예로, 환시 생성 모델은 다운 샘플링(down-sampling) 합성곱 레이어, 업 샘플링(up-sampling) 합성곱 레이어, 잔차(residual) 합성곱 레이어, 및 활성화 함수 중의 적어도 하나를 포함할 수 있다.

[0048] 장치는 특정 이미지 및 환시 이미지를 깊이 추정 모델에 입력하여 특정 이미지에 대한 제1 깊이 맵 및 환시 이미지에 대한 제2 깊이 맵을 획득할 수 있다(S120).

[0049] 깊이 추정 모델은 깊이 추정 인코딩부 및 깊이 추정 디코딩부를 포함할 수 있다. 일 예로, 장치는 특정 이미지 및 환시 이미지를 깊이 추정 인코딩부에 입력하여 제1 인코딩 특징맵을 획득할 수 있다. 그리고, 장치는 제1 인코딩 특징맵 및 사용자의 키 대비 깊이 비율을 깊이 추정 디코딩부에 입력하여 제1 깊이 맵 및 제2 깊이 맵을 획득할 수 있다.

[0050] 장치는 제1 깊이 맵 및 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성할 수 있다(S130).

[0051] 장치는 특정 이미지, 환시 이미지, 제1 깊이 맵, 제2 깊이 맵, 및 제1 깊이 볼륨 데이터에 기초하여 점유 벡터를 획득할 수 있다(S140).

[0052] 구체적으로, 장치는 특정 이미지, 환시 이미지, 제1 깊이 맵 및 제2 깊이 맵을 2D 기하 인코딩부에 입력하여 제2 인코딩 특징맵을 획득할 수 있다. 장치는 제1 깊이 볼륨 데이터를 3D 기하 인코딩부에 입력하여 제3 인코딩 특징맵을 획득할 수 있다.

[0053] 그리고, 장치는 제2 인코딩 특징 맵 및 제3 인코딩 특징 맵을 점유 추정 모델에 입력하여 점유 벡터를 획득할 수 있다. 여기서, 점유 벡터는 점유 여부를 알고자 하는 특정 점(x)에 대한 점유 정보를 나타내는 벡터를 의미한다.

[0054] 장치는 점유 벡터에 기초하여 획득된 3D 메시(mesh) 정보를 통해 사용자에 대한 3D 인간 모델을 복원할 수 있다(S150).

[0055] 구체적으로, 장치는 점유 벡터에 대해 마칭 큐브(marching cube) 알고리즘을 적용하여 3D 메시 정보가 획득할 수 있다.

[0056] 그리고, 장치는 특정 이미지, 환시 이미지, 제1 깊이 맵, 제2 깊이 맵, 및 제2 깊이 볼륨 데이터에 기초하여, 색상 예측 정보를 획득할 수 있다. 장치는 색상 예측 정보 및 3D 메시 정보에 기초하여 사용자에 대한 3D 인간 모델을 복원할 수 있다.

[0057] 여기서, 제2 깊이 볼륨 데이터는 특정 이미지를 3차원 공간에 투영한 후 이에 대해 픽셀 값을 투영하여 획득된 깊이 볼륨 데이터를 의미할 수 있다.

[0058] 한편, 장치는 환시 추정 모델, 깊이 추정 모델 및 점유 추정 모델 각각을 학습시키기 위한 손실 함수의 값을 산출할 수 있다. 환시 추정 모델, 깊이 추정 모델 및 점유 추정 모델 각각을 학습시키기 위한 손실 함수에 입력될 변수 데이터는 데이터 셋 내에 수집될 수 있다.

[0059] 즉, 장치는 데이터 셋 내에서 변수를 선택한 후 이를 각 손실 함수에 입력하여 환시 추정 모델, 깊이 추정 모델

및 점유 추정 모델 각각을 학습시키기 위한 손실 함수의 값을 산출할 수 있다.

[0060] 장치는 산출된 손실 함수의 값을 환시 추정 모델, 깊이 추정 모델 및 점유 추정 모델 각각에 대해 역전파(back-propagation)함으로써 환시 추정 모델, 깊이 추정 모델 및 점유 추정 모델 각각을 학습시킬 수 있다.

[0062] 도 2는 본 개시의 일 실시예에 따른, 딥러닝 기반의 인간 3D 모델 복원 신경망 파이프라인을 설명하기 위한 도면이다.

[0063] 도 2에 도시된 바와 같이, 인간 3D 모델 복원 신경망 파이프라인은 데이터 셋(즉, 데이터 셋 구성부), 딥러닝 네트워크 1-환시 생성기, 딥러닝 네트워크 2-깊이 추정 인코딩 부, 딥러닝 네트워크 3-크기 추정부, 딥러닝 네트워크 4-깊이 추정 디코딩 부, 깊이 투영 부, 마칭 큐브 알고리즘(Marching Cubes Algorithm), 딥러닝 네트워크 5-2D 기하 인코딩 부, 딥러닝 네트워크 6-3D 기하 인코딩 부, 딥러닝 네트워크 7-모델 점유 추정 부, 색상 투영 부, 딥러닝 네트워크 8-2D 색상 인코딩 부, 딥러닝 네트워크 9-3D 색상 인코딩 부, 딥러닝 네트워크 10-모델 색상 추정 부, 손실 함수 계산부 1, 손실 함수 계산부 2, 손실 함수 계산부 3, 손실 함수 계산부 4, 환시 생성기 학습부, 깊이 추정기 학습부, 기하 신경망 학습부, 색상 신경망 학습부를 포함할 수 있다.

[0064] 본 개시를 설명함에 있어서, "모듈"은 신경망 파이프라인을 이루는 분업화된 객체를 의미한다.

[0065] "인코딩"은 입력 영상을 컨볼루션 신경망을 통해 "인코딩 특징 맵"을 추출하는 행위를 의미한다. "인코딩 특징 맵"은 다른 모듈의 입력 데이터로서 입력될 수 있다.

[0066] "디코딩"은 인코딩 특징 맵 및 인코딩 부 컨볼루션 레이어(Layer) 각 출력으로부터의 스킵 커넥션(skip connection)을 U-Net 구조로 입력 받음으로써 컨볼루션 신경망을 통해 "디코딩 특징 맵"을 추출하는 행위를 의미한다.

[0067] "환시(Hallucination)"는 실제 3D 구조로부터 렌더링된 영상이 아니라, 신경망에 그럴듯한 조건을 부여하여 자연스러운 복원을 수행하기 위해 환시 생성기에 의해 만들어지는 RGB 3 채널의 영상을 의미한다.

[0068] "점유(Occupancy)"는 3D 좌표계 상에서 주어진 점이 모델의 표면(Surface)의 안인지 밖인지를 판별할 수 있게 하는 값이다.

[0069] 일 예로, 3D 좌표계 상에서 주어진 점이 모델의 표면의 밖일 경우, 점유에 대응되는 수치는 0으로 표현될 수 있다. 3D 좌표계 상에서 주어진 점이 모델의 표면의 안일 경우, 점유에 대응되는 수치는 1로 표현될 수 있다. 3D 좌표계 상에서 주어진 점이 모델의 표면에 대응될 경우, 점유에 대응되는 수치는 0.5로 표현될 수 있다.

[0070] "다운 샘플링(Downsampling)"은 합성곱 계층이 입력 텐서(Tensor)의 크기를 축소시키며 특징 맵을 추출하는 연산을 의미한다.

[0071] "업 샘플링(Upsampling)"은 전치 합성곱(Transposed Convolution) 계층이 입력 텐서의 크기를 확대하며 출력 맵을 만들어내는 연산을 의미한다.

[0072] "잔차(Residual)" 연산은 특정 신경망 계층이 출력을 내보낼 때, 입력된 텐서를 더해주는 연산을 의미한다.

[0073] "스킵 커넥션(Skip-Connection)"은 U-Net 구조의 신경망에서, 인코더에 있는 각 레이어의 출력을 디코더의 각 레이어의 입력으로 추가적으로 더해주는 동작을 의미한다.

[0074] "U-Net"은 인코더-디코더의 구조를 가지면서, 인코더 각 레이어의 출력을 디코더의 각 레이어에 추가적으로 연결하여, 인코더의 높은 레벨 특징 맵(High-Level Feature Map) 및 입력 영상의 정보를 동시에 활용할 수 있도록 만든 신경망 구조이다.

[0075] 이하에서는 인간 3D 모델 복원 신경망 파이프라인에 포함된 모듈의 동작을 설명하도록 한다.

[0076] "데이터 셋 구성부"는 딥러닝 네트워크를 학습시키기 위한 데이터 셋 제작 과정을 수행할 수 있다. 일 예로, "데이터 셋 구성부"는 3D 모델의 스캐닝, 렌더링, 3D 모델의 키 대비 깊이 값에 의한 깊이 맵 보정 등을 수행할 수 있다.

[0077] "딥러닝 네트워크 1-환시 생성기"는 단일 입력 영상 I^F를 활용하여 입력 영상의 뒷면 환시 I^B를 생성(즉, $I^B=F_H(I^F)$)할 수 있다.

[0078] 본 개시의 일 예로, 도 2에 도시된 바와 같이, 환시 생성기는 다운샘플링 합성곱 계층, 업샘플링 합성곱 계층, 잔차 합성곱 계층 및 활성화 함수로 구성될 수 있다.

[0079] 그리고, 본 개시의 일 예로, 도 3에 도시된 바와 같이, 환시 생성기를 구성하는 잔차 합성곱 모듈은 그룹 정규화 계층, 합성곱 계층, 활성화 함수 등을 포함할 수 있다.

[0080] "손실 함수 계산부 1"는 환시 생성기 학습을 위하여 손실 값을 산출할 수 있다. "손실 함수 계산부 1"는 사실적인 뒷모습을 생성하기 위하여 3 가지 손실 함수(예로, 스무스(smoothed L1) 손실 함수(L_1), 인지 손실(perceptual loss) 함수(L_p), 패치 판별기(patch discriminator) f_{PD}를 활용한 적대적(adversarial) 손실 함수(L_{adv}))를 사용할 수 있다.

[0081] 구체적으로, smoothed L1 손실 함수는 수학식 1과 같이 구현될 수 있다. 수학식 1에서 E는 픽셀 계산 값의 평균 연산을 의미할 수 있다.

수학식 1

[0082]
$$\mathcal{L}_1 = \begin{cases} \mathbb{E}[0.5(\hat{I}^B - I^B)^2] & \text{if } |\hat{I}^B - I^B| < 1, \\ \mathbb{E}[|\hat{I}^B - I^B| - 0.5] & \text{otherwise,} \end{cases}$$

[0083] 인지 함수 함수는 수학식 2와 같이 구현될 수 있다. 수학식 2에서 Φ는 미리 학습된 VGG-19 모델을 의미할 수 있다.

수학식 2

[0084]
$$\mathcal{L}_p = \mathbb{E}[\sum_{j=1}^{n} |\phi_j(\hat{I}^B) - \phi_j(I^B)|],$$

[0085] 적대적 손실 함수는 수학식 3과 같이 구현될 수 있다. 수학식 3에서 a, b, 및 c는 f_{PD}를 학습/훈련시키기 위한 라벨(label) 맵을 의미하며, $f_{PD}(x)$의 출력 텐서와 크기가 같을 수 있다. 일 예로, b 및 c의 모든 값은 1일 수 있으며, a는 0일 수 있다.

수학식 3

[0086]
$$\begin{aligned} \min_{f_{PD}} \mathcal{L}_{adv}(f_{PD}) &= \frac{1}{2}\mathbb{E}[(f_{PD}(\hat{I}^B) - b)^2] \\ &\quad + \frac{1}{2}\mathbb{E}[(f_{PD}(f_H(I^F)) - a)^2], \\ \min_{f_H} \mathcal{L}_{adv}(f_H) &= \frac{1}{2}\mathbb{E}[(f_{PD}(f_H(I^F)) - c)^2], \end{aligned}$$

[0087] "환시 생성기 학습부"는 L_1 + L_p + L_{adv} 결과 값을 환시 생성기 신경망 내에서 역전파(backpropagation)시킴으로써 환시 생성기를 학습시킬 수 있다.

[0088] 상술한 바와 같이, 환시 생성기는 적대적 손실 함수를 포함한 다양한 훈련 방법을 통해 인지적으로 그럴듯한 뒷면 영상을 생성할 수 있다. 아무 조건 없이 완전 연결 신경망만을 사용하여 가려진 부분에 대해 추정을 시도하는 것이 아닌, 공간적인 정보를 처리하는데 효과적인 합성곱 신경망을 이용하여 만든 환시로 완전 연결 신경망의 추정을 도와줄 수 있다.

[0089] "딥러닝 네트워크 2-깊이 추정 인코딩 부"는 모델의 앞뒤 영상(즉, I^F 및 I^B)을 입력 받아 인코딩 특징 맵(F_d)을 출력할 수 있다.

[0090] "딥러닝 네트워크 3-크기 추정부"는 인코딩 특징 맵(F_d)를 입력 받아 모델의 키 대비 깊이 비율인 λ를 추정할 수 있다.

[0091] "딥러닝 네트워크 4-크기 추정 디코딩 부"는 F_d를 입력 받아 깊이 맵 $\overline{D}^F$ 및 $\overline{D}^B$를 도출할 수 있다. "딥러닝 네트워크 4-크기 추정 디코딩 부"는 깊이 맵 $\overline{D}^F$ 및 $\overline{D}^B$ 각각에 λ를 곱하여 깊이 추정기 최종 출력인 D^F 및 D^B 각각을 획득할 수 있다.

[0092] 상술한 바와 같이, 최종 출력인 D^F 및 D^B를 출력하는 깊이 추정기(즉, 깊이 추정 네트워크)는 도 5와 같이 구성될 수 있다.

[0093] 도 5에서 f_s는 SMPL 복셀 인코더를 의미할 수 있다. 추정된 SMPL 볼륨(Vo)을 통해 SMPL 복셀 특징이 출력될 수 있다. 복셀 인코더는 2D 깊이 방향 컨볼루션 레이어를 포함할 수 있다. 복셀 인코더에 의해 3D 볼륨의 각 깊이 레이어에 2D 컨볼루션이 적용될 수 있다. 이에 따라, 깊이 레이어 별로 Vo를 인코딩될 수 있다.

[0094] 그리고, fD_{enc}는 I^F 및 환각된 뒷면 이미지 I^B를 입력으로 하여 특징을 추출할 수 있다. fD_{dec}는 F_S 및 F_d에 기초하여 깊이를 예측할 수 있다.

[0095] 본 개시의 일 예로, 도 6은 도 5에서 λ(예로, 모델의 키 대비 깊이 값)(즉, 스케일 팩터)를 획득하는 과정을 도시하는 도면이다.

[0096] "손실 함수 계산부 2"는 깊이 추정기(즉, 깊이 추정부)를 학습시키기 위한 손실 함수 L_{req} 및 L_{scale}을 산출할 수 있다.

[0097] L_{req}는 수학식 4와 같이 구현될 수 있으며, L_{scale}는 수학식 5와 같이 구현될 수 있다. 수학식 5에서 $\hat{\lambda}$는 정답 깊이 맵의 최댓값을 의미할 수 있다.

수학식 4

[0098] $$\mathcal{L}_{reg} = \mathbb{E}[(\log\hat{D} - \log D)],$$

수학식 5

[0099] $$\mathcal{L}_{scale} = \mathbb{E}[(\hat{\lambda} - \lambda)^2].$$

[0100] 전체 깊이 손실 값은 수학식 6에 따라 산출될 수 있다.

수학식 6

[0101] $$\mathcal{L}_{depth} = \mathcal{L}_{reg} + \eta\mathcal{L}_{scale},$$

[0102] "깊이 추정기 학습부"는 L_{depth} 결과 값을 깊이 추정기 신경망에 역전파시킴으로써 깊이 추정기(또는, 크기 추정부)를 학습시킬 수 있다.

[0103] "깊이 투영부"는 D^F 및 D^B를 3차원 공간 상에 투영할 수 있다. D^F,D^B 각각으로 만들어지는 텐서를 V^F,V^B라 할 때, V^F 및 V^B의 합집합 V(즉, 최종 투영 텐서)가 산출될 수 있다.

[0104] V^F 및 V^B 각각은 앞면 깊이 맵을 투영하여 생성된 텐서 및 뒷면 깊이 맵을 투영하여 생성된 텐서를 의미할 수

있다. V^F 및 V^B 각각은 수학식 7 및 수학식 8에 따라 구현될 수 있다.

수학식 7

[0105] $$V^F_{i,j,k} = \begin{cases} 1 & \text{if } k = R(D^F(i,j) + \psi^F), \\ 0 & \text{otherwise,} \end{cases}$$

수학식 8

[0106] $$V^B_{i,j,k} = \begin{cases} 1 & \text{if } k = R(1 - (D^B(i,j) + \psi^B)), \\ 0 & \text{otherwise,} \end{cases}$$

[0107] 이 때, 깊이 맵 투영을 V의 가운데로 모아주기 위한 오프셋은 앞뒤 오프셋(ψ^F), 글로벌 오프셋(ψ^G), 보상 오프셋(ψ^F, ψ^F), 및 노이즈 오프셋(ψ^F)을 포함할 수 있다.

[0108] 앞뒤 오프셋(ψ^F)은 수학식 9와 같이 구현될 수 있으며, 글로벌 오프셋(ψ^G)은 수학식 10과 같이 구현될 수 있으며, 보상 오프셋(ψ^F, ψ^F)은 수학식 11와 같이 구현될 수 있으며, 노이즈 오프셋(ψ^F)은 수학식 12와 같이 구현될 수 있다.

수학식 9

[0109] $$\psi^F = \psi_G - \psi^F_C, \quad \psi^B = \psi_G - \psi^B_C,$$

수학식 10

[0110] $$\psi_G = \frac{1}{2}(1 - \max(D^F_{max}, D^B_{max})).$$

수학식 11

[0111] $$\psi^F_C = \begin{cases} \psi_N & \text{if } D^F_{max} \geq D^B_{max}, \\ \frac{1}{2}|D^F_{max} - D^B_{max}| & \text{otherwise,} \end{cases}$$
$$\psi^B_C = \begin{cases} \psi_N & \text{if } D^F_{max} < D^B_{max}, \\ \frac{1}{2}|D^F_{max} - D^B_{max}| & \text{otherwise,} \end{cases}$$

수학식 12

[0112] $$\psi_N = \begin{cases} f_{clip}(f_N(-\frac{1}{2}|D^F_{max} - D^B_{max}|, \frac{1}{2}|D^F_{max} - D^B_{max}|)), & \textit{if training} \\ 0, & \textit{otherwise.} \end{cases}$$

[0114] 상술한 바와 같이, 깊이 추정기를 통해 깊이를 추정하고 이를 적절한 투영 알고리즘에 기초하여 투영함으로써,

완전 연결 모듈이 추정하기 이전에 기하 정보 및 이의 복셀에 해당하는 색상이 생성될 수 있다.

[0115] 기하 추정을 위한 투영 센서는 V^G로 칭할 수 있다. 그리고, 색상 추정을 위한 투영 텐서의 경우, "색상 투영부"에서 입력 RGB 영상을 투영함으로써 수학식 7 및 8에서 1 대신 픽셀 값을 투영한 텐서를 V^C로 칭할 수 있다.

[0116] "딥러닝 네트워크 5-2D 기하 인코딩부"는 는 I^F, I^B, D^F, D^B를 특징 맵 F_{2d}으로 인코딩하는 합성곱 모듈이다.

[0117] "딥러닝 네트워크 6-3D 기하 인코딩 부"는 V^G를 특징 맵 F_{3d}으로 인코딩하는 합성곱 모듈이다.

[0118] 점유 여부를 알고 싶은 점 x에 대해, "딥러닝 네트워크 7-모델 점유 추정 부"는 F_{2d}를 x에 대해 UV 2D 투영하고 F_{3d}를 x에 대해 3D 투영한 뒤, 이를 완전 연결 계층에 입력하여 x의 점유 여부를 판별할 수 있다. 이 때, "딥러닝 네트워크 7-모델 점유 추정 부"가 출력한 점유 벡터(Occupancy Vector)를 p라 한다.

[0119] 모델 메쉬(mesh)를 획득하기 위하여, 점유 벡터에 마칭 큐브 알고리즘이 적용될 수 있다.

[0120] "손실 함수 계산부 3"은 정답 3D 스캔에서 추출한 점유 벡터 $\hat{p}$를 비교하여 점유 손실 값을 획득할 수 있다. 점유 손실 값은 수학식 13와 같이 구현될 수 있다.

수학식 13

[0121]
$$\mathcal{L}_{occ} = \frac{1}{n}\sum_{i=1}^{n}((\hat{p}_i - p_i)^2).$$

[0122] "기하 신경망 학습부"는 $\mathcal{L}_{occ}$의 결과값을 기하 추정 신경망에 역전파함으로써 기하 추정 신경망을 학습시킬 수 있다.

[0123] "딥러닝 네트워크 8-2D 색상 인코딩 부"는 I^F, I^B, D^F, D^B를 특징 맵 F_{2d}으로 인코딩하는 합성곱 모듈이다.

[0124] "딥러닝 네트워크 9-3D 색상 인코딩 부"는 V^C를 특징 맵 F_{3d}으로 인코딩하는 합성곱 모듈이다.

[0125] 색상을 추정하고자 하는 점 x에 대해, "딥러닝 네트워크 10-모델 색상 추정 부"는 F_{2d}를 x에 대해 UV 2D 투영하고 F_{3d}를 x에 대해 3D 투영한 뒤, 이를 완전 연결 계층에 입력하여 x의 색상을 추정할 수 있다. 이 때, "딥러닝 네트워크 10-모델 색상 추정 부"는 5채널의 RGB $C(p)_t$(3) 및 마스크 $\alpha\beta$(2) 채널 벡터를 출력할 수 있다.

[0126] 모델의 최종 색상은 입력 영상을 활용하는 부분 및 이를 최종 색상으로 더하는 부분으로 구성될 수 있다.

[0127] 입력 영상을 활용하는 부분은 수학식 14(즉, 앞뒤 입력 영상 조합식)와 같이 구현될 수 있으며, 이를 최종 색상으로 더하는 부분은 수학식 15와 같이 구현될 수 있다.

수학식 14

[0128]
$$C(p)_\beta = \beta X(I^F, \pi(x)) + (1-\beta)X(I^B, \pi(x))$$

수학식 15

[0129] $$C(p) = \alpha C(p)_{\beta} + (1-\alpha)C(p)_{t}.$$

[0130] 수학식 14에서 π(x)는 3 차원 점 x에 대한 2D 투영 결과(UV 투영 맵)이며, X는 π(x)에 해당하는 샘플링을 위한 쌍선형 샘플링 보간법(bilinear sampling interpolation)에 대응될 수 있다.

[0131] "손실 함수 계산부 4"는 색상 추정 신경망의 훈련을 위한 손실(즉, 컬러 손실 및 알파 손실)을 획득할 수 있다. 컬러 손실은 수학식 16와 같이 구현될 수 있고, 알파 손실은 수학식 17와 같이 구현될 수 있다.

수학식 16

[0132] $$\mathcal{L}_{clr} = \frac{1}{n}\sum_{i=1}^{n}(|\hat{C}(p_i) - C(p_i)| + |\hat{C}(p_i) - C(p_i)_t|),$$

수학식 17

[0133] $$\mathcal{L}_{\alpha} = -\mathbb{E}[\log(\alpha)].$$

[0134] 전체 텍스쳐 손실은 컬러 손실 및 알파 손실의 합(L_{tex})으로 산출될 수 있다.

[0135] "색상 신경망 학습부"는 L_{tex}의 결과 값을 색상 추정 신경망에 역전파함으로써 색상 추정 신경망을 학습시킬 수 있다.

[0136] 종래의 단일 영상 입력 네트워크 중 완전 연결 신경망 기반의 알고리즘들은 공통적으로 크게 두 가지 한계를 가질 수 있다.

[0137] 첫째로, 종래의 기술이 적용될 경우, 모델 뒷면의 정보를 전혀 알 수 없어 기하 및 색상의 추정이 자연스럽지 않을 수 있다.

[0138] 둘째로, 종래의 기술이 적용될 경우, 합성곱 신경망이 갖는 귀납 편향이라는 가정을 사용할 수 없다.

[0139] 여기서, 완전 연결 신경망이 Mean Squared Error(MSE) 손실 함수를 최소화하기 위하여 단일 입력에서 보이지 않는 부분에 대해 스무딩(예로, 기하 관련 정보에 대해 스무딩) 및 블러링(예로, 색상 관련 정보에 대해 블러링)시킨 후 이에 대해 회색으로 추정하는 경향성을 가진다.

[0140] 입력받은 사용자의 실제 뒷모습과 본 개시에서 제시하는 환시 이미지에서 나타내는 실제 생김새가 다를 수 있다.

[0141] 다만, 본 개시의 환시 생성기는 적대적 손실 함수를 포함한 다양한 훈련 방법을 통해 인지적으로 그럴듯한 뒷면 영상을 생성할 수 있다. 이는, 아무 조건 없이 완전 연결 신경망만을 사용하여 가려진 부분에 대한 추정을 방식과 비교할 때, 공간적인 정보를 처리하는 데 적합한 합성곱 신경망을 이용해 만든 환시 이미지로 완전 연결 신경망의 추정을 도와주는 방식은 높은 성능을 보인다.

[0142] 뿐만 아니라, 본 개시는 깊이 추정기를 통해 깊이를 추정하고, 이를 적절한 투영 알고리즘으로 투영함으로써 완전 연결 모듈이 추정하기 이전에 기하와 그 복셀에 해당하는 색상을 추정/생성할 수 있다.

[0143] 이러한 방식으로, 본 개시는 완전 연결 추정 모듈이 인간 복원이라는 문제를 해결할 때 최대한 쉬운 조건에서 해결할 수 있도록 유도하였다.

[0145] 도 7은 본 개시의 실시예에 따른 장치를 예시하는 블록도이다.

[0146] 도 7을 기준하면, 장치(100)는 본 개시에서 설명된 3D 인간 모델을 복원하는 장치를 나타낼 수 있다.

[0147] 예로, 장치(100)는 환시를 통해 생성된 인간의 뒷모습에 대한 깊이 추정을 통해 인간의 3D 모델을 생성할 수 있다.

[0148] 장치(100)는 프로세서(110), 메모리(120), 송수신부(130), 입력 인터페이스 장치(140) 및 출력 인터페이스 장치(150) 중 적어도 하나를 포함할 수 있다. 각각의 구성 요소들은 공통 버스(bus)(160)에 의해 연결되어 서로 통신을 수행할 수 있다. 또한, 각각의 구성 요소들은 공통 버스(160)가 아니라, 프로세서(110)를 중심으로 개별 인터페이스 또는 개별 버스를 통하여 연결될 수도 있다.

[0149] 프로세서(110)는 AP(Application Processor), CPU(Central Processing Unit), GPU(Graphic Processing Unit) 등과 같은 다양한 종류들로 구현될 수 있으며, 메모리(120)에 저장된 명령을 실행하는 임의의 반도체 장치일 수 있다. 프로세서(110)는 메모리(120)에 저장된 프로그램 명령(program command)을 실행할 수 있다. 프로세서(110)는 상술한 도 1 내지 도 6을 토대로 설명한 움직이는 카메라로 촬영한 비디오 영상에 기초하여 프레임별로 깊이 맵을 추출하는 방법을 수행할 수 있다.

[0150] 프로세서(110)는 환시를 통해 생성된 인간의 뒷모습에 대한 깊이 추정을 통해 인간의 3D 모델을 생성할 수 있다.

[0151] 그리고/또는, 프로세서(110)는 하나 이상의 모듈들에 대한 적어도 하나의 기능을 구현하기 위한 프로그램 명령을 메모리(120)에 저장하여, 도 1 내지 도 6에 기반하여 설명된 동작이 수행되도록 제어할 수 있다. 즉, 도 1 내지 도 6에 따른 각 동작 및/또는 기능은 하나 이상의 프로세서(110)에 의해 실행될 수 있다.

[0152] 메모리(120)는 다양한 형태의 휘발성 또는 비 휘발성 저장 매체를 포함할 수 있다. 예를 들어, 메모리(120)는 ROM(read-only memory) 및 RAM(random access memory)을 포함할 수 있다. 본 개시의 실시예에서 메모리(120)는 프로세서(110)의 내부 또는 외부에 위치할 수 있고, 메모리(120)는 이미 알려진 다양한 수단을 통해 프로세서(110)와 연결될 수 있다.

[0153] 예를 들어, 메모리(120)는 사용자에 대한 이미지, 환시 이미지, 깊이 맵 등을 저장할 수 있다.

[0154] 송수신부(130)는 프로세서(110)에서 처리된/처리될 데이터를 외부 장치 및/또는 외부 시스템과 송수신하는 기능을 수행할 수 있다.

[0155] 예를 들어, 송수신부(130)는 다른 단말 장치와 데이터 교환 등에 활용될 수 있다.

[0156] 입력 인터페이스 장치(140)는 데이터를 프로세서(110)로 제공하도록 구성된다.

[0157] 출력 인터페이스 장치(150)는 프로세서(110)로부터의 데이터를 출력하도록 구성된다.

[0159] 본 개시의 예시적 실시예들에서 설명된 구성요소는, 하드웨어 요소에 의해 구현될 수 있다. 예를 들어, 상기 하드웨어 요소는, 디지털 신호 프로세서(DSP), 프로세서, 컨트롤러, 주문형 집적 회로(ASIC), FPGA와 같은 프로그래머블 로직 요소, GPU, 기타 전자 장치, 또는 이들의 조합 중 적어도 하나를 포함할 수 있다. 본 개시의 예시적 실시예들에서 설명된 기능 또는 과정 중 적어도 일부는 소프트웨어로 구현될 수 있으며, 소프트웨어는 기록 매체에 기록될 수 있다. 예시적 실시예들에서 설명하는 구성요소, 기능 및 프로세스는 하드웨어와 소프트웨어의 조합으로 구현될 수 있다.

[0160] 본 개시의 일 실시예에 따른 방법은 컴퓨터에 의해 수행될 수 있는 프로그램으로 구현될 수 있으며, 상기 컴퓨터 프로그램은 자기 저장 매체, 광 판독 매체, 디지털 저장 매체 등과 같은 다양한 기록 매체에 기록될 수 있다.

[0161] 본 개시에서 설명한 다양한 기술들은, 디지털 전자 회로 또는 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합으로 구현될 수 있다. 상기 기술들은 컴퓨터 프로그램 제품, 즉, 정보 매체에 유형적으로 구현되는 컴퓨터 프로그램 또는 컴퓨터 프로그램(예를 들어, 기계 판독 가능 저장 장치(예: 컴퓨터 판독 가능 매체) 또는 데이터 처리 장치), 데이터 처리 장치에 의해 처리되거나 데이터 처리 장치(예를 들어, 프로그램 가능한 프로세서, 컴퓨터 또는 복수의 컴퓨터들)의 동작을 하기 위해 전파된 신호로 구현되는 컴퓨터 프로그램으로 구현될 수 있다.

[0162] 컴퓨터 프로그램(들)은 컴파일된 언어나 해석된 언어를 포함한 모든 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형 프로그램 또는 모듈, 구성 요소, 서브루틴 또는 컴퓨팅 환경에서 사용하기에 적합한 기타 단위를 포함하는 모든 형태로 배포될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터, 또는 한 사이트 또는 여러 사이트에 분산되어 있고 통신 네트워크로 상호 연결되어 있는 복수 컴퓨터들에 의해 수행될 수 있다.

[0163] 컴퓨터 프로그램의 실행에 적합한 프로세서의 예로, 범용 및 특수 목적 마이크로프로세서, 및 디지털 컴퓨터의 하나 이상의 프로세서가 포함된다. 일반적으로 프로세서는 읽기 전용 메모리나 랜덤 액세스 메모리 또는 둘 모두에서 명령과 데이터를 수신한다. 컴퓨터의 구성 요소는 명령을 실행하기 위한 적어도 하나의 프로세서, 및 명령 및 데이터를 저장하기 위한 하나 이상의 메모리 장치를 포함할 수 있다. 또한, 컴퓨터는 데이터 저장을 위한 하나 이상의 대용량 저장 장치, 예를 들어, 자기, 광자기 디스크 또는 광 디스크를 포함하거나, 상기 대용량 저장 장치에 연결되어 데이터를 수신 및/또는 전송을 수행할 수 있다. 컴퓨터 프로그램 명령 및 데이터를 구현하기에 적합한 정보 매체의 예로는, 반도체 메모리 장치(예를 들어, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체), 컴팩트 디스크 읽기 전용 메모리(CD- ROM), 디지털 비디오 디스크(DVD) 등과 같은 광학 매체, 플롭티컬 디스크와 같은 광자기 매체, 및 ROM(Read Only Memory), RAM(Random Access Memory), 플래시 메모리, EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 및 기타 알려진 컴퓨터 판독 가능 매체가 포함된다. 프로세서와 메모리는 특수 목적 논리 회로에 의해 보완되거나 통합될 수 있다.

[0164] 프로세서는 운영 체제(OS) 및 OS에서 실행되는 하나 이상의 소프트웨어 애플리케이션을 실행할 수 있다. 프로세서 장치는 또한 소프트웨어 실행에 응답하여 데이터를/에 액세스, 저장, 조작, 처리 및 생성할 수 있다. 단순화를 위해 프로세서 장치는 단수로 설명되었으나, 해당 기술 분야의 통상의 지식을 가진자는, 프로세서 장치가 복수의 처리 요소들 및/또는 다양한 타입의 처리 요소들을 포함할 수 있음을 이해할 수 있다. 예를 들어, 프로세서 장치는 복수 프로세서들 또는 프로세서(processor)와 제어부(controller)를 포함할 수 있다. 또한, 병렬 프로세서들과 같이 상이한 처리 구조를 구성할 수도 있다. 또한, 컴퓨터가 읽을 수 있는 매체는 컴퓨터가 접근할 수 있는 모든 매체를 의미하며, 컴퓨터 저장매체와 전송매체를 모두 포함할 수 있다.

[0165] 본 개시는 다양한 상세 구현 예들의 세부 설명을 포함하나, 그 세부 사항이 본 개시에서 제안하는 발명 또는 청구 범위를 제한하는 것은 아니고, 특정한 예시적 실시예의 특징을 설명하는 것으로 이해되어야 한다.

[0166] 본 개시에서 예시적 실시예들에서 개별적으로 설명된 특징들은 단일 예시적 실시예에 의해 구현될 수 있다. 반대로, 본 개시에서 단일 예시적 실시예에 대해 설명된 다양한 특징들은 복수의 예시적 실시예들의 조합 또는 적절한 하위 조합에 의해 구현될 수도 있다. 나아가, 본 개시에서, 상기 특징들은 특정 조합에 의해 동작할 수 있고, 최초 상기 조합이 청구된 것처럼 기술될 수 있으나, 경우에 따라, 하나 이상의 특징이 청구된 조합에서 제외되거나, 청구된 조합이 하위 조합 또는 하위 조합의 수정 형태로 변경될 수 있다.

[0167] 마찬가지로, 도면에 특정 순서로 동작이 설명되어 있다 하더라도, 목표된 결과를 획득하기 위해, 동작들을 특정 순번 또는 순서로 실행하는 것이 필요하거나, 모든 동작들의 수행이 필요한 것으로 이해되어서는 안 된다. 특정한 경우에는 멀티태스킹과 병렬 처리가 유용할 수 있다. 또한, 모든 실시예들의 예시적 실시예들에서 다양한 장치 구성요소들이 반드시 분리되어야 하는 것으로 이해되어서는 안 되며, 전술한 프로그램 구성요소 및 장치는 단일 소프트웨어 제품 또는 여러 소프트웨어 제품에 패키징될 수 있다.

[0168] 본 명세서에 개시된 예시적 실시예들은 단지 예시적인 것이며 본 개시의 범위를 제한하려는 것이 아니다. 해당 기술 분야의 통상의 지식을 가진자는, 청구범위 및 그 균등물의 사상 및 범위를 벗어나지 않으면서 예시적 실시예들에 대한 다양한 수정이 이루어질 수 있음을 인식할 수 있을 것이다.

[0169] 따라서, 본 개시는 이하의 특허청구범위 내에 속하는 모든 다른 교체, 수정 및 변경을 포함한다고 할 것이다.

부호의 설명

[0170] 100 : 장치 110 : 프로세서

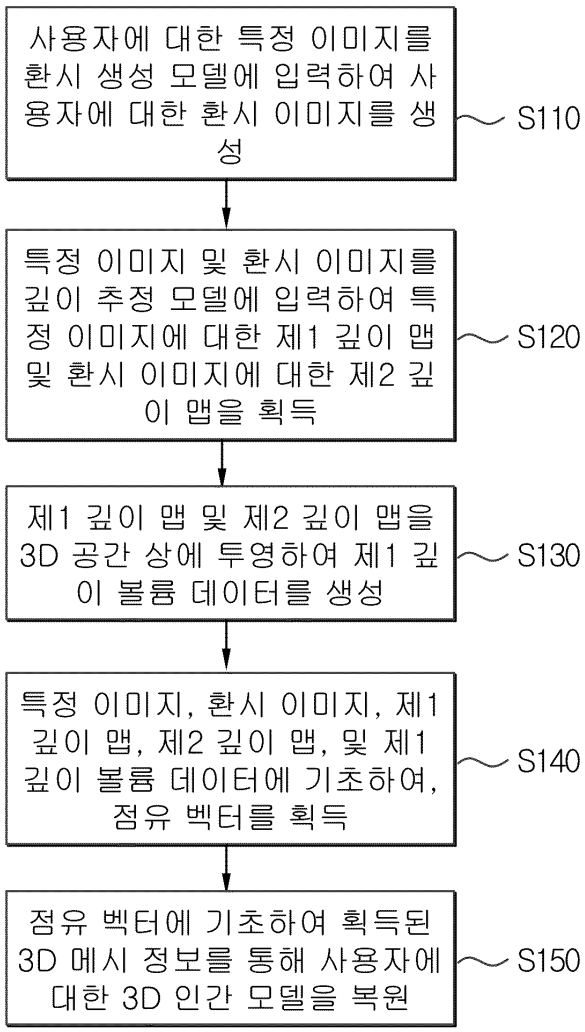
120 : 메모리 130 : 송수신부

140 : 입력 인터페이스 장치 150 : 출력 인터페이스 장치

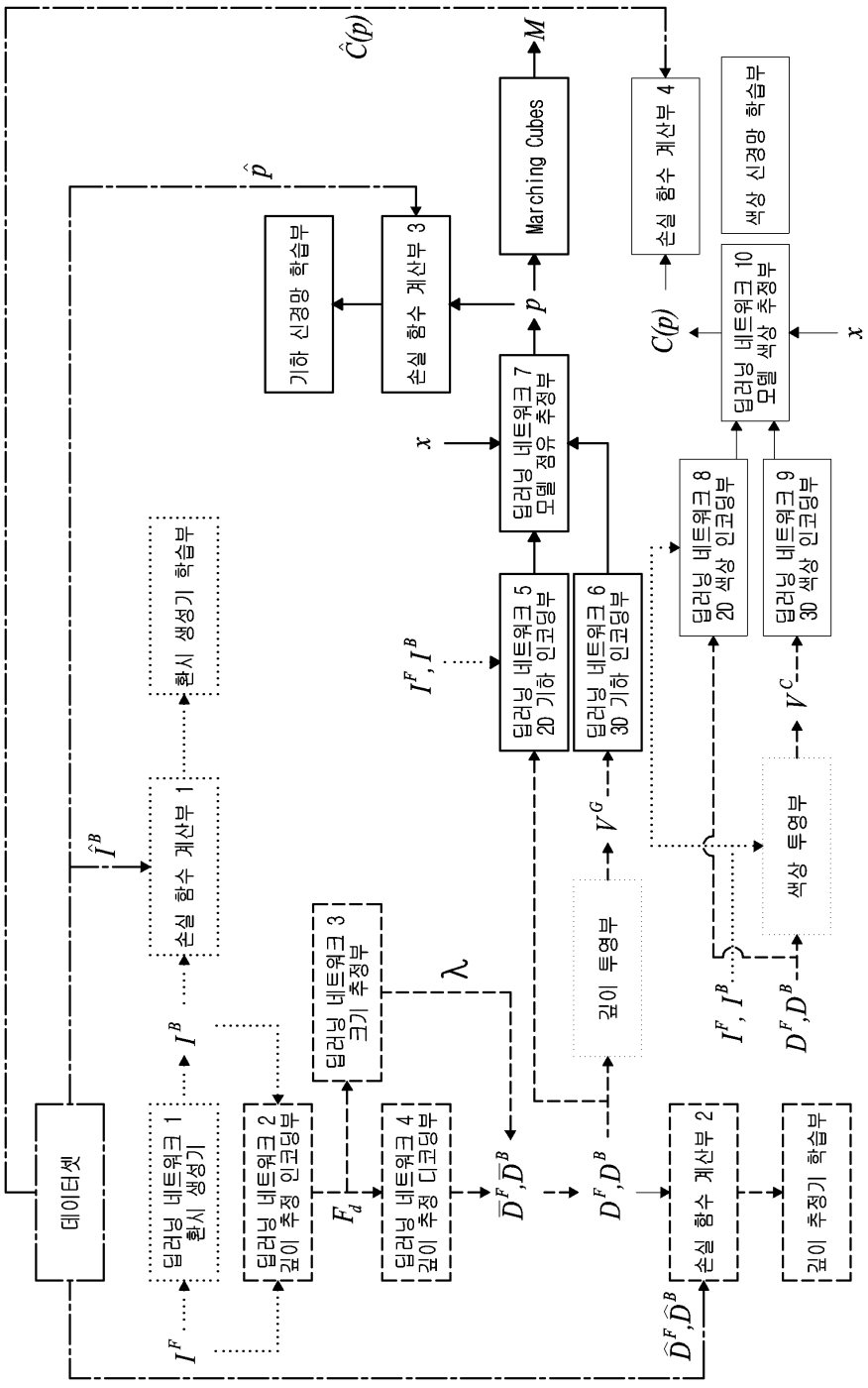
160 : 공통 버스

도면

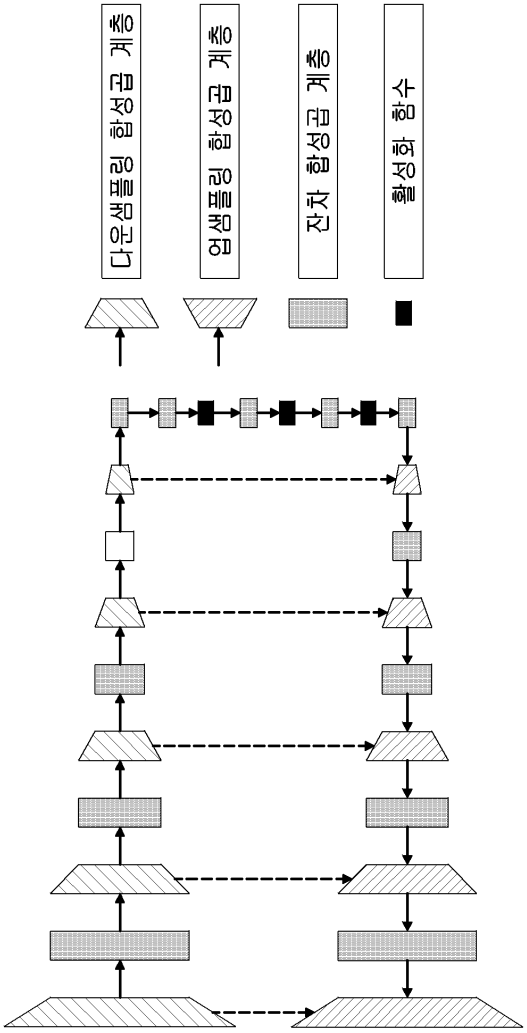
도면1


사용자에 대한 특정 이미지를 환시 생성 모델에 입력하여 사용자에 대한 환시 이미지를 생성
S110
특정 이미지 및 환시 이미지를 깊이 추정 모델에 입력하여 특정 이미지에 대한 제1 깊이 맵 및 환시 이미지에 대한 제2 깊이 맵을 획득
S120
제1 깊이 맵 및 제2 깊이 맵을 3D 공간 상에 투영하여 제1 깊이 볼륨 데이터를 생성
S130
특정 이미지, 환시 이미지, 제1 깊이 맵, 제2 깊이 맵, 및 제1 깊이 볼륨 데이터에 기초하여, 점유 벡터를 획득
S140
점유 벡터에 기초하여 획득된 3D 메시 정보를 통해 사용자에 대한 3D 인간 모델을 복원
S150

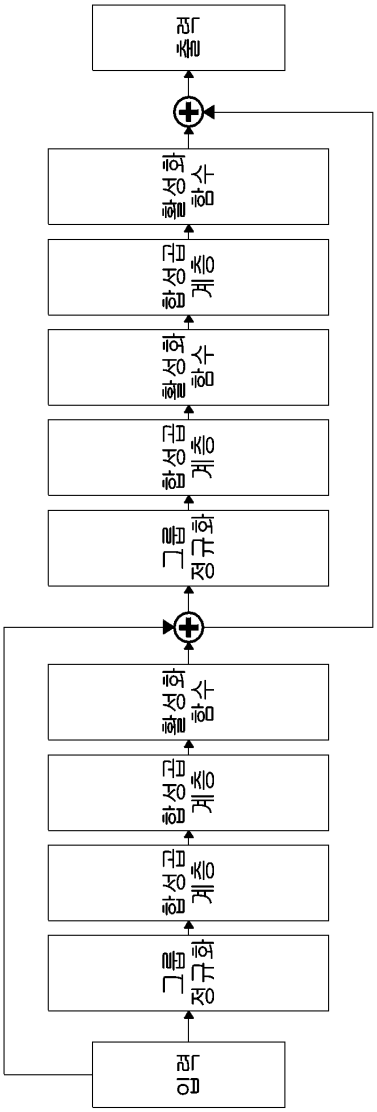
도면2


데이터셋
딥러닝 네트워크 1 환시 생성기
딥러닝 네트워크 2 깊이 추정 인코딩부
딥러닝 네트워크 3 크기 추정부
딥러닝 네트워크 4 깊이 추정 디코딩부
손실 함수 계산부 1
환시 생성기 학습부
깊이 추정기 학습부
손실 함수 계산부 2
깊이 투영부
딥러닝 네트워크 5 2D 기하 인코딩부
딥러닝 네트워크 6 3D 기하 인코딩부
딥러닝 네트워크 7 모델 점유 추정부
기하 신경망 학습부
손실 함수 계산부 3
Marching Cubes
색상 투영부
딥러닝 네트워크 8 2D 색상 인코딩부
딥러닝 네트워크 9 3D 색상 인코딩부
딥러닝 네트워크 10 모델 색상 추정부
손실 함수 계산부 4
색상 신경망 학습부

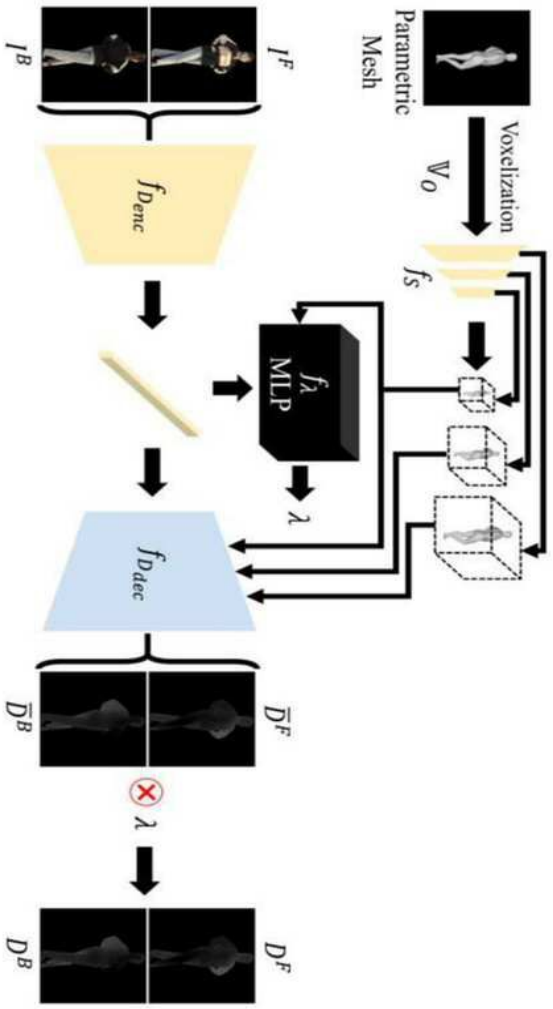
도면3



도면4



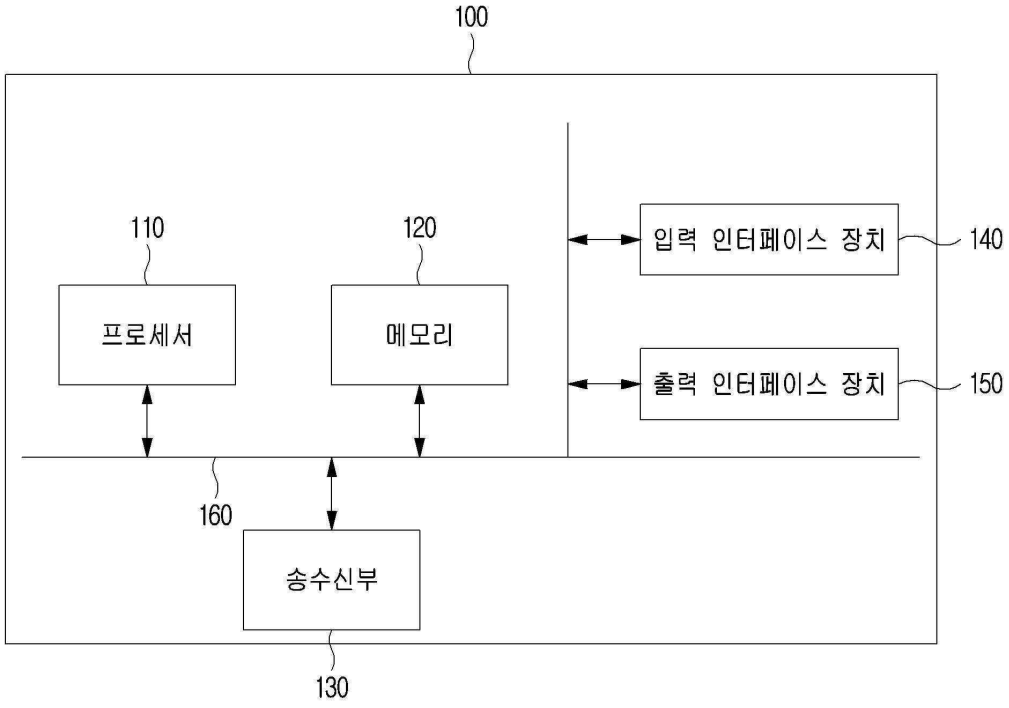
도면5


Parametric Mesh
Voxelization
V_O
f_S
I^B
I^F
$f_{D_{enc}}$
f_λ MLP
λ
$f_{D_{dec}}$
$\bar{D}^B$
$\bar{D}^F$
λ
D^B
D^F

도면6


ZY Ratio $\hat{\lambda} = \frac{Z_{max}}{Y_{max}}$
Y
Y_{max}
Z_{max}
Z

도면7


100
110
프로세서
120
메모리
입력 인터페이스 장치
140
출력 인터페이스 장치
150
160
송수신부
130