(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0152042
(43) 공개일자 2024년10월21일

(51) 국제특허분류(Int. Cl.)
G06T 17/20 (2006.01) *G06T 13/40* (2011.01)
G06T 19/20 (2011.01) *G06T 7/593* (2017.01)
G06T 7/90 (2017.01)
(52) CPC특허분류
G06T 17/20 (2013.01)
G06T 13/40 (2013.01)
(21) 출원번호 10-2023-0048114
(22) 출원일자 2023년04월12일
심사청구일자 2023년04월12일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
박인규
서울특별시 강남구 선릉로 221, 103동 1403호(도곡동, 도곡렉슬아파트)
정태현
서울특별시 동작구 동작대로29길 91, 206동 1106호 (사당동, 사당우성아파트)
(74) 대리인
양성보

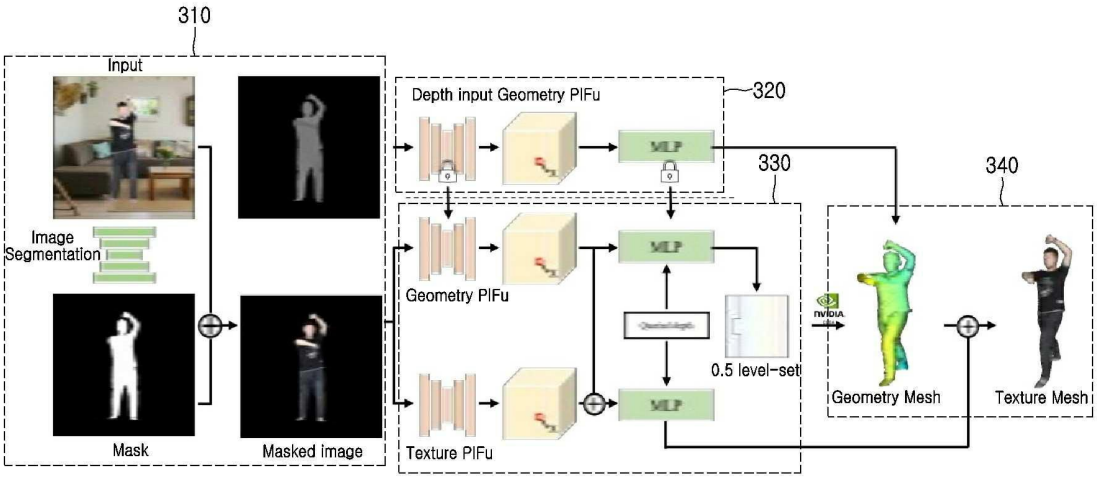
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **영상으로부터 3차원 휴먼 형상 복원 방법 및 시스템**

(57) 요 약

영상으로부터 3차원 휴먼 형상 복원 방법 및 시스템이 제시된다. 본 발명에서 제안하는 영상으로부터 3차원 휴먼 형상 복원 방법은 그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계, 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계, 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 단계 및 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06T 19/20 (2013.01)
G06T 7/593 (2017.01)
G06T 7/90 (2017.01)
G06T 2207/10028 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711179347
과제번호 00155915
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신인재양성(교육운영수익)
연구과제명 [Ezbaro][정부+민간] 인공지능융합혁신인재양성(2차년도)
기 여 율 10/100
과제수행기관명 인하대학교 산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711160508
과제번호 2022-0-00981-001
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 방송통신산업기술개발사업
연구과제명 [Ezbaro] 전배경 정합 3D 객체 스트리밍 기술개발(2차년도)
기 여 율 90/100
과제수행기관명 주식회사 로커스
연구기간 2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계;

상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계;

상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 단계;

상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계

를 포함하는 3차원 휴먼 메쉬 형상 복원 방법.

청구항 2

제1항에 있어서,

상기 그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계는,

다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 툴을 통해 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세를 포함하는 데이터셋으로 생성하고, 특징점의 3차원 위치를 입력 받아 3차원 메쉬의 표면 주변에서 정규 분포를 따라 배치하는

3차원 휴먼 메쉬 형상 복원 방법.

청구항 3

제1항에 있어서,

상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계는,

학습 데이터셋을 입력 받아 특징 맵을 추출하고, 3차원 공간의 차폐 정보를 구하기 위한 특징 점의 2차원 위치를 상기 특징 맵에 직교 투영하여 상기 특징 맵에서의 위치 정보 값을 양선형 보간법을 통해 구하고, 투영된 특징맵은 깊이 정보와 함께 딥-음함수에 입력되어 상기 특징 점에서의 딥-음함수 결과를 출력하는

3차원 휴먼 메쉬 형상 복원 방법.

청구항 4

제3항에 있어서,

상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계는,

상기 딥-음함수 추론 네트워크에서 깊이 영상에 대하여 유클리디안 손실 함수를 통해 학습하고, 상기 환각 네트워크에서 상기 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 학습하며, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 최적화를 위해 상기 특징 맵의 유클리디안 차이를 이용하는

3차원 휴먼 메쉬 형상 복원 방법.

청구항 5

그래픽스 툴을 통한 학습 데이터셋을 생성하는 학습 데이터 셋 생성부;

상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 깊이 영상 학습부;

상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 RGB 영상 학습부; 및

상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 3차원 휴먼 메쉬 형상 복원부

를 포함하는 3차원 휴먼 메쉬 형상 복원 시스템.

청구항 6

제5항에 있어서,

상기 학습 데이터 셋 생성부는,

다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 툴을 통해 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세를 포함하는 데이터셋으로 생성하고, 특징점의 3차원 위치를 입력 받아 3차원 메쉬의 표면 주변에서 정규 분포를 따라 배치하는

3차원 휴먼 메쉬 형상 복원 시스템.

청구항 7

제5항에 있어서,

상기 깊이 영상 학습부는,

학습 데이터셋을 입력 받아 특징 맵을 추출하고, 3차원 공간의 차폐 정보를 구하기 위한 특징 점의 2차원 위치를 상기 특징 맵에 직교 투영하여 상기 특징 맵에서의 위치 정보 값을 양선형 보간법을 통해 구하고, 투영된 특징맵은 깊이 정보와 함께 딥-음함수에 입력되어 상기 특징 점에서의 딥-음함수 결과를 출력하는

3차원 휴먼 메쉬 형상 복원 시스템.

청구항 8

제7항에 있어서,

상기 3차원 휴먼 메쉬 형상 복원부는,

상기 딥-음함수 추론 네트워크에서 깊이 영상에 대하여 유클리디안 손실 함수를 통해 학습하고, 상기 환각 네트워크에서 상기 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 학습하며, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 최적화를 위해 상기 특징 맵의 유클리디안 차이를 이용하는

3차원 휴먼 메쉬 형상 복원 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 영상으로부터 3차원 휴먼 형상 복원 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 증강현실의 시스템적 구현과 증강 객체 복원 기술이 큰 관심을 받고있다. 특히 고성능의 3차원 휴먼 형상 복원 기술은 메타버스 플랫폼에서 K-pop, 1인 미디어의 발전과 함께 큰 미래 가치가 예상되는 연구이다. 기존의 3차원 형상 복원 연구는 스튜디오 환경에서 촬영한 다중뷰 영상에서의 기하학적 정합을 통해 구현되었다. 하지만 이는 특정 촬영 환경에서 취득한 영상에서만 활용이 가능하며, 일반적인 사용자가 3차원 형상 복원을 수행하는 것은 불가능했다. 그러나 최근 심층신경망 기술의 발전과 함께 단안 영상에서 3차원 객체 형상을 복원하는 연구가 크게 발전하고 있으며, 특히 객체를 사람으로만 한정하여 복원할 수 있는 3차원 휴먼 형상 복원 연구가 크게

주목 받고 있다.

[0003] 이러한 3차원 휴먼 형상 복원 연구는 크게 변수 기반의 방식과 비변수 복원 방식으로 나뉜다. 변수 복원 방식의 경우 SMPL, SMPL-X와 같이 마네킹 형식의 완성된 휴먼 형상을 통해 객체의 모양과 자세를 추정한다. 이러한 기술은 완벽한 형상의 객체 복원이 가능하다는 점에서 큰 장점을 가지지만 마네킹과 같은 결과를 출력하여 머리카락, 드레스 등 일반적인 형상은 복원할 수 없다는 단점을 가진다. 이러한 일반적인 형상을 복원하기 위해 변수 기반 데이터셋을 추론에 직접 활용하지 않는 비변수 기반 방식이 크게 주목 받고 있다. 고전적인 비변수 기반 모델은 3차원 공간에 대한 복셀을 추론하여 형상 복원을 구현한다. 하지만 해당 방식은 3차원 복셀 추론으로 인해 많은 처리 속도를 요구하거나, 변수 데이터셋을 부분적으로 활용하여 낮은 성능 결과를 출력한다. 이를 해결하기 위해 딥-음함수와 픽셀 위치 추론 처리를 휴먼 형상 복원에 적용하여 모든 문제의 해결할 수 있는 기술 PIFu(Pixel-aligned Function)이 제시되었다. PIFu는 RGB 영상을 CNN(Convolutional Neural Network) 구조에 입력하여 특징맵을 추론하고, 원하는 위치의 3차원 위치 정보를 특징맵에 입력하고 MLP(Multilayer perceptron)을 통해 각 점의 내부/외부 정보를 추론하는 네트워크이다. 최근에는 이를 활용한 여러 양상의 연구가 진행되고 있다. 학습 과정에 SMPL 추론 네트워크를 추가해 공간적 모호성을 해결한 PaMIR, 휴먼 객체 분할의 전처리 과정과 특정 카메라 시점에서의 2차원 영상만을 복원하여 실시간 구동을 구현하는 등 다양한 연구 방향성이 존재한다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) [1] S. Saito, Z. Huang, R. Natsume, S. Morishima, H. Li, and A. Kanazawa. PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization. In Proc. of IEEE International Conference on Computer Vision, pages 2304-2314, November 2019.

(특허문헌 0002) [2] J. Hoffman, S. Gupta, and T. Darrell. Learning with side information through modality hallucination. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2016.

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적 과제는 비변수 기반 방식의 신경망 학습 과정에서 추가적인 공간 정보인 휴먼 깊이 영상을 이용하여 객체의 안정적인 형상을 복원하며, 실시간 구동이 가능한 3차원 휴먼 형상 복원 방법 및 시스템을 제공하는데 있다. 제안하는 3차원 휴먼 형상 복원 방법 및 시스템은 휴먼 깊이 영상 데이터 셋 구축하고, 깊이 영상 기반의 딥-음함수 네트워크 학습한 후, 깊이 영상 네트워크를 활용한 RGB 입력 영상에서의 지식 증류의 세 단계로 나눠 진다.

과제의 해결 수단

[0006] 일 측면에 있어서, 본 발명에서 제안하는 영상으로부터 3차원 휴먼 형상 복원 방법은 그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계, 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계, 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 단계 및 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계를 포함한다.

[0007] 상기 그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계는 다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 툴을 통해 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세를 포함하는 데이터셋으로 생성하고, 특징점의 3차원 위치를 입력 받아 3차원 메쉬의 표면 주변에서 정규 분포를 따라 배치한다.

[0008] 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계는 학습 데이터셋을 입력 받아 특징 맵을 추출하고, 3차원 공간의

차폐 정보를 구하기 위한 특징 점의 2차원 위치를 상기 특징 맵에 직교 투영하여 상기 특징 맵에서의 위치 정보 값을 양선형 보간법을 통해 구하고, 투영된 특징맵은 깊이 정보와 함께 딥-음함수에 입력되어 상기 특징 점에서의 딥-음함수 결과를 출력한다.

[0009] 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계는 상기 딥-음함수 추론 네트워크에서 깊이 영상에 대하여 유클리디안 손실 함수를 통해 학습하고, 상기 환각 네트워크에서 상기 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 학습하며, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 최적화를 위해 상기 특징 맵의 유클리디안 차이를 이용한다.

[0010] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 영상으로부터 3차원 휴먼 형상 복원 시스템은 그래픽스 툴을 통한 학습 데이터셋을 생성하는 학습 데이터 셋 생성부, 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 깊이 영상 학습부, 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 RGB 영상 학습부 및 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 3차원 휴먼 메쉬 형상 복원부를 포함한다.

발명의 효과

[0011] 본 발명의 실시예들에 따르면 3차원 휴먼 형상 복원 방법 및 시스템을 통해 휴먼 깊이 영상 데이터 셋 구축하고, 깊이 영상 기반의 딥-음함수 네트워크 학습한 후, 깊이 영상 네트워크를 활용한 RGB 입력 영상에서의 지식 증류를 수행하며, 비변수 기반 방식의 신경망 학습 과정에서 추가적인 공간 정보인 휴먼 깊이 영상을 이용하여 객체의 안정적인 형상을 실시간으로 복원할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 3차원 휴먼 형상 복원 방법을 설명하기 위한 흐름도이다.

도 2는 본 발명의 일 실시예에 따른 OpenGL 기반의 3차원 휴먼 학습을 위한 데이터셋 생성 과정을 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 실시간 동적 3차원 휴먼 형상 복원 과정을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 3차원 휴먼 형상 복원 시스템의 구성을 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 실시간 동적 3차원 휴먼 객체 검출 및 분할 예시를 나타내는 도면이다.

도 6은 본 발명의 일 실시예에 따른 사전 학습된 깊이 영상 모델로부터 지식 증류를 통한 안정적인 3차원 휴먼 형상 복원 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0013] 본 발명은 머리카락, 드레스 등 일반적인 사람 형상을 복원할 수 있는 비변수 기반의 효율적인 3차원 휴먼 형상 복원 방식을 제안한다. 본 발명은 사전 학습된 깊이 영상 입력 네트워크로부터 RGB 입력 네트워크로의 지식 증류 과정을 통한 안정 적인 휴먼 3차원 형상 복원 기술을 구현한다. 또한 데이터 후처리 과정에 GPGPU 가속화된 마칭 큐브(Marching cube) 알고리즘을 사용하여 실시간 구동이 가능한 3 차원 휴먼 메쉬 형상 출력을 목표한다. 제안하는 방식은 그래픽스 툴 OpenGL을 통한 학습 데이터셋 생성, 사전 학습된 깊이 영상 모델에서의 지식 증류 기반 딥-음함수 추론 네트워크 학습, CUDA 기반의 GPGPU 가속화된 3차원 메쉬 형상 복원의 세 단계를 포함한다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0015] 도 1은 본 발명의 일 실시예에 따른 3차원 휴먼 형상 복원 방법을 설명하기 위한 흐름도이다.

[0016] 제안하는 3차원 휴먼 형상 복원 방법은 그래픽스 툴을 통한 학습 데이터셋을 생성하는 단계(110), 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행하는 단계(120), 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행하는 단계(130) 및 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워

크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원하는 단계(140)를 포함한다.

[0017] 단계(110)에서, 그래픽스 툴을 통한 학습 데이터셋을 생성한다.

[0018] 본 발명의 실시예에 따르면, 다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 툴을 통해 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세를 포함하는 데이터셋으로 생성하고, 특징점의 3차원 위치를 입력 받아 3차원 메쉬의 표면 주변에서 정규 분포를 따라 배치한다.

[0019] 단계(120)에서, 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행한다.

[0020] 본 발명의 실시예에 따르면, 학습 데이터셋을 입력 받아 특징 맵을 추출하고, 3차원 공간의 차폐 정보를 구하기 위한 특징 점의 2차원 위치를 상기 특징 맵에 직교 투영하여 상기 특징 맵에서의 위치 정보 값을 양선형 보간법을 통해 구하고, 투영된 특징맵은 깊이 정보와 함께 딥-음함수에 입력되어 상기 특징 점에서의 딥-음함수 결과를 출력한다.

[0021] 본 발명의 실시예에 따른 딥-음함수 네트워크 학습은 기존 RGB 네트워크와 첫 번째 레이어를 제외하고 모두 동일한 구조를 가지지만, 깊이 영상에서 3차원 정보를 직접 입력 받으므로 매우 높은 성능의 복원 결과를 출력한다. 본 발명에서는 효율적인 학습 성능 증가 확인을 위하여 전면 깊이 영상 입력 외에도 전, 후면 깊이 영상을 함께 입력 받는 네트워크에서의 학습을 진행하며 이에 대한 형상 복원 성능을 비교한다.

[0022] 단계(130)에서, 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행한다.

[0023] 본 발명의 실시예에 따르면, 사전 학습된 깊이 입력 모델로부터 RGB 영상을 입력으로 받는 환각 네트워크를 증류 학습한다. 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크는 영상을 입력 받는 첫 번째 레이어를 제외하고는 모두 동일한 구조를 가진다.

[0024] 단계(140)에서, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원한다.

[0025] 상기 딥-음함수 추론 네트워크에서 깊이 영상에 대하여 유클리디안 손실 함수를 통해 학습하고, 상기 환각 네트워크에서 상기 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 학습하며, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 최적화를 위해 상기 특징 맵의 유클리디안 차이를 이용한다.

[0026] 깊이 영상 모델로부터 환각 네트워크로의 지식 증류를 위해 네트워크 간 ReLU, leaky ReLU, sigmoid의 모든 활성 함수 레이어에서의 손실함수를 계산하고 이를 학습에 활용한다. 활성 함수의 입력값에 sigmoid 값을 취하고 이를 통해 네트워크를 최적화한다. 이러한 학습 방식을 통해 본 발명에서는 동일한 네트워크에서의 안정적인 시스템 모델 구동을 구현한다. 각 점의 RGB 값을 출력하는 텍스쳐 추론 네트워크는 휴먼 형상 복원 네트워크와 유사하게 ResNet 인코더와 MLP가 합쳐진 구조이며, 인코더의 특징 맵은 형상 복원 인코더의 특징 맵과 합쳐져 MLP에 입력된다. 출력 결과는 3채널의 RGB 값을 추론하도록 학습된다. 실제 구동 시에는 깊이 영상 네트워크를 제외한 RGB 입력의 환각 네트워크만으로 0.5 레벨 집합 추론을 수행하며 텍스쳐 추론 네트워크, 후처리 알고리즘과 함께 3차원 휴먼 메쉬 형상을 복원한다.

[0028] 도 2는 본 발명의 일 실시예에 따른 OpenGL 기반의 3차원 휴먼 학습을 위한 데이터셋 생성 과정을 설명하기 위한 도면이다.

[0029] 3차원 휴먼 형상 복원을 위해서는 사람 형상의 정확한 공간적 위치 정보를 가지는 학습 데이터셋이 요구된다. 본 발명의 실시예에서는 다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 API OpenGL을 통해 가공 및 제작하여 사용한다. 생성한 가상 데이터는 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세로 구성된다. 형상 복원 네트워크는 임의의 조명 환경에서도 높은 성능을 기록해야 하며 사람의 촬영 방향과 관계없이 안정적으로 형상 복원을 수행해야 한다. 따라서 데이터셋의 다양성을 위해 무작위로 조명을 변경함과 동시에 휴먼 메쉬 형상을 1씩 360 회전시켜 도 2와 같이 하나의 휴먼 데이터(210)에서 무작위 조명 환경에서 취득한 360개의 데이터셋(220)을 구축한다. 딥-음함수는 3차원 공간에 존재하는 연속적인

0.5 레벨 집합 추정 함수로, 특징점의 3차원 위치를 입력 받아 물체의 표면을 기준으로 내부는 1, 외부는 0 값을 출력한다. 본 발명에서는 딥-음함수의 효율적 학습을 위하여 3차원 공간에 균일하거나 무작위로 배치하지 않고 3차원 메쉬의 표면 주변에서 N(0, σ)의 정규 분포를 따라 배치하여 효율적인 학습이 가능한 데이터셋을 제작한다.

[0031] 도 3은 본 발명의 일 실시예에 따른 실시간 동적 3차원 휴먼 형상 복원 과정을 설명하기 위한 도면이다.

[0032] 본 발명의 실시예에 따르면, 단일 사람이 존재하는 RGB 입력 영상에서 머리카락, 드레스 등 일반적인 휴먼 형상 복원이 가능한 비변수 기반의 실시간 구동 시스템 구현을 목표한다.

[0033] 제안하는 실시간 동적 3차원 휴먼 형상 복원 과정은 입력 동영상의 모든 프레임에서 동작하며 휴먼 객체 검출 및 분할, 3차원 딥-음함수 추론, GPGPU 가속화된 Marching cube 알고리즘을 통한 완성된 3차원 휴먼 형상 복원의 세 단계로 나눠 진다.

[0034] 먼저, 휴먼 객체 검출 및 분할 과정(Image Segmentation)을 수행한다(310). 3 차원 휴먼 형상 복원 시 배경이 있는 영상을 입력 받는다면 입력 영상이 모호해지므로 형상 복원이 어렵다. 이에 과분할 과정을 추가하여 안정적인 휴먼 형상 획득을 구현한다. 분할된 휴먼 영상은 PIFu의 딥-음함수 네트워크로 입력 되어 3차원 0.5 레벨 집합과 각 특징점에 대한 RGB 값을 추론할 특징맵을 추론한다(320). 본 발명의 실시예에서는 이때 안정적인 형상 복원을 위해 사전 학습된 깊이 영상 모델로부터 지식 증류 학습을 수행한다. 마지막으로 0.5 레벨 집합에서 Marching cube 및 RGB 추론을 GPGPU 고속 구동(330)하여 완성된 3차원 휴먼 메쉬 형상을 복원한다(340).

[0035] 본 발명의 실시예에 따른 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습에 대하여 설명한다.

[0036] 딥-음함수 추론 네트워크 학습은 3차원 공간의 차폐 정보를 표현하는 0.5 레벨 집합 표현 음함수 f를 구해 3차원 형상을 추론한다. 함수 f는 식(1)과 같이 3차원 공간 내, 점 위치 $P = (P_x, P_y, P_z) \in \mathbb{R}^3$와 입력 영상 I를 입력 받으며, 만약 해당 점 P가 물체 표면을 경계로 내부 위치 시 1을, 외부 위치 시 0을 출력한다. 즉, 딥-음함수 f는 0.5 레벨 집합 추론 함수를 의미하며 물체 표면을 기준으로 3차원 공간에 대한 실내, 외 정보를 나타낸다.

$$f(P,I) = \begin{cases} 1 & \text{if P inside mesh surface} \\ 0 & \text{otherwise} \end{cases}$$

[0037] (1)

[0038] 본 발명에서는 이러한 딥-음함수 추론 방식에 픽셀 정합 개념을 도입한 비변수 기반 3차원 휴먼 형상 모델 PIFu를 제안한다. 딥-음함수 f는 식(2)의 CNN 구조 특징 맵 출력부 Φ와 MLP 구조 차폐 공간 추론부 g로 나뉜다. 특징맵 출력부는 휴먼 영상 I를 입력 받아 특징 맵을 추출한다.

[0039] $$f(P,I) = g(\pi(P_x, P_y, \Phi(I)), P_z) \quad (2)$$

[0040] 이후 차폐 정보를 구할 점 P의 2차원 위치 P_x, P_y를 특징 맵에 직교 투영하여 특징 맵에서의 위치 정보 값 $\pi(P_x, P_y, \Phi(I))$를 양선형 보간법을 통해 구한다. 투영된 특징맵은 깊이 정보인 P_z와 함께 함수 g에 입력되어 P에서의 딥-음함수 결과, 즉 차폐 정보를 출력한다. 이러한 투영 방식을 통해 3차원 처리가 아닌 2차원 추론 과정으로 축소하여 효율적인 구동을 구현한다. PIFu는 학습 시에는 메쉬 주변의 점을 무작위 추출하여 최적화를 진행하며, 실제 구동 시에는 3차원 격자점의 모든 차폐 정보를 계산한 후, 마칭 큐브를 통해 3차원 형상을 복원한다.

[0042] 도 4는 본 발명의 일 실시예에 따른 3차원 휴먼 형상 복원 시스템의 구성을 나타내는 도면이다.

[0043] 본 실시예에 따른 3차원 휴먼 형상 복원 시스템(400)은 프로세서(410), 버스(420), 네트워크 인터페이스(430), 메모리(440) 및 데이터베이스(450)를 포함할 수 있다. 메모리(440)는 운영체제(441) 및 3차원 휴먼 형상 복원

루틴(442)을 포함할 수 있다. 프로세서(410)는 학습 데이터 셋 생성부(411), 깊이 영상 학습부(412), RGB 영상 학습부(413) 및 3차원 휴먼 메쉬 형상 복원부(414)를 포함할 수 있다. 다른 실시예들에서 3차원 휴먼 형상 복원 시스템(400)은 도 4의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 3차원 휴먼 형상 복원 시스템(400)은 디스플레이나 트랜시버(transceiver)와 같은 다른 구성요소들을 포함할 수도 있다.

[0044] 메모리(440)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(440)에는 운영체제(441)와 3차원 휴먼 형상 복원 루틴(442)을 위한 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 드라이브 메커니즘(drive mechanism, 미도시)을 이용하여 메모리(440)와는 별도의 컴퓨터에서 판독 가능한 기록 매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록 매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록 매체(미도시)를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록 매체가 아닌 네트워크 인터페이스(430)를 통해 메모리(440)에 로딩될 수도 있다.

[0045] 버스(420)는 3차원 휴먼 형상 복원 시스템(400)의 구성요소들간의 통신 및 데이터 전송을 가능하게 할 수 있다. 버스(420)는 고속 시리얼 버스(high-speed serial bus), 병렬 버스(parallel bus), SAN(Storage Area Network) 및/또는 다른 적절한 통신 기술을 이용하여 구성될 수 있다.

[0046] 네트워크 인터페이스(430)는 3차원 휴먼 형상 복원 시스템(400)을 컴퓨터 네트워크에 연결하기 위한 컴퓨터 하드웨어 구성요소일 수 있다. 네트워크 인터페이스(430)는 3차원 휴먼 형상 복원 시스템(400)을 무선 또는 유선 커넥션을 통해 컴퓨터 네트워크에 연결시킬 수 있다.

[0047] 데이터베이스(450)는 3차원 휴먼 형상 복원을 위해 필요한 모든 정보를 저장 및 유지하는 역할을 할 수 있다. 도 4에서는 3차원 휴먼 형상 복원 시스템(400)의 내부에 데이터베이스(450)를 구축하여 포함하는 것으로 도시하고 있으나, 이에 한정되는 것은 아니며 시스템 구현 방식이나 환경 등에 따라 생략될 수 있고 혹은 전체 또는 일부의 데이터베이스가 별개의 다른 시스템 상에 구축된 외부 데이터베이스로서 존재하는 것 또한 가능하다.

[0048] 프로세서(410)는 기본적인 산술, 로직 및 3차원 휴먼 형상 복원 시스템(400)의 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(440) 또는 네트워크 인터페이스(430)에 의해, 그리고 버스(420)를 통해 프로세서(410)로 제공될 수 있다. 프로세서(410)는 학습 데이터 셋 생성부(411), 깊이 영상 학습부(412), RGB 영상 학습부(413) 및 3차원 휴먼 메쉬 형상 복원부(414)를 위한 프로그램 코드를 실행하도록 구성될 수 있다. 이러한 프로그램 코드는 메모리(440)와 같은 기록 장치에 저장될 수 있다.

[0049] 학습 데이터 셋 생성부(411), 깊이 영상 학습부(412), RGB 영상 학습부(413) 및 3차원 휴먼 메쉬 형상 복원부(414)는 도 1의 단계들(110~140)을 수행하기 위해 구성될 수 있다.

[0050] 3차원 휴먼 형상 복원 시스템(400)은 학습 데이터 셋 생성부(411), 깊이 영상 학습부(412), RGB 영상 학습부(413) 및 3차원 휴먼 메쉬 형상 복원부(414)를 포함할 수 있다.

[0051] 학습 데이터 셋 생성부(411)는 그래픽스 툴을 통한 학습 데이터셋을 생성한다.

[0052] 본 발명의 실시예에 따른 학습 데이터 셋 생성부(411)는 다중뷰 스튜디오 환경에서 취득한 3차원 메쉬 휴먼 데이터를 그래픽스 툴을 통해 RGB 영상, 깊이 영상, 객체 분할 영상, UV 맵, 딥-음함수의 결과 값, 카메라 자세를 포함하는 데이터셋으로 생성하고, 특징점의 3차원 위치를 입력 받아 3차원 메쉬의 표면 주변에서 정규 분포를 따라 배치한다.

[0053] 깊이 영상 학습부(412)는 상기 생성된 학습 데이터셋을 이용하여 사전 학습된 깊이 영상 모델에서의 형상 추론을 위한 지식 증류 기반 딥-음함수 추론 네트워크 학습을 수행한다.

[0054] 본 발명의 실시예에 따른 깊이 영상 학습부(412)는 학습 데이터셋을 입력 받아 특징 맵을 추출하고, 3차원 공간의 차폐 정보를 구하기 위한 특징 점의 2차원 위치를 상기 특징 맵에 직교 투영하여 상기 특징 맵에서의 위치 정보 값을 양선형 보간법을 통해 구하고, 투영된 특징맵은 깊이 정보와 함께 딥-음함수에 입력되어 상기 특징점에서의 딥-음함수 결과를 출력한다.

[0055] 본 발명의 실시예에 따른 딥-음함수 네트워크 학습은 기존 RGB 네트워크와 첫 번째 레이어를 제외하고 모두 동일한 구조를 가지지만, 깊이 영상에서 3차원 정보를 직접 입력 받으므로 매우 높은 성능의 복원 결과를 출력한다. 본 발명에서는 효율적인 학습 성능 증가 확인을 위하여 전면 깊이 영상 입력 외에도 전, 후면 깊이 영상을

함께 입력 받는 네트워크에서의 학습을 진행하며 이에 대한 형상 복원 성능을 비교한다.

[0056] RGB 영상 학습부(413)는 상기 사전 학습된 사전 학습된 깊이 영상 모델로부터 RGB 영상을 입력으로 받아 환각 네트워크 증류 학습을 수행한다.

[0057] 본 발명의 실시예에 따른 RGB 영상 학습부(413)는 사전 학습된 깊이 입력 모델로부터 RGB 영상을 입력으로 받는 환각 네트워크를 증류 학습한다. 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크는 영상을 입력 받는 첫 번째 레이어를 제외하고는 모두 동일한 구조를 가진다.

[0058] 3차원 휴먼 메쉬 형상 복원부(414)는 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 레이어에서의 손실함수를 계산하고 학습에 반영하여 3차원 휴먼 메쉬 형상을 복원한다.

[0059] 상기 딥-음함수 추론 네트워크에서 깊이 영상에 대하여 유클리디안 손실 함수를 통해 학습하고, 상기 환각 네트워크에서 상기 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 학습하며, 상기 딥-음함수 추론 네트워크 및 상기 환각 네트워크 간의 모든 활성 함수 최적화를 위해 상기 특징 맵의 유클리디안 차이를 이용한다.

[0061] 도 5는 본 발명의 일 실시예에 따른 실시간 동적 3차원 휴먼 객체 검출 및 분할 예시를 나타내는 도면이다.

[0062] 도 5에 도시된 바와 같이 휴먼 객체 검출 및 분할 과정을 수행한다. 도 5(a)와 같이 입력 영상을 입력 받아, 도 5(b)와 같이 객체를 검출한다. 3 차원 휴먼 형상 복원 시 배경이 있는 영상을 입력 받는다면 입력 영상이 모호해지므로 형상 복원이 어렵다. 이후, 분할 과정(도 5(c))을 추가하여 안정적인 휴먼 형상 획득을 구현한다. 분할된 휴먼 영상은 PIFu의 딥-음함수 네트워크로 입력 되어 3차원 0.5 레벨 집합과 각 특징점에 대한 RGB 값을 추론할 특징맵을 추론한다. 본 발명의 실시예에서는 이때 안정적인 형상 복원을 위해 사전 학습된 깊이 영상 모델로부터 지식 증류 학습을 수행한다. 마지막으로 0.5 레벨 집합에서 Marching cube 및 RGB 추론을 GPGPU 고속 구동하여 완성된 3차원 휴먼 메쉬 형상을 복원한다.

[0064] 도 6은 본 발명의 일 실시예에 따른 사전 학습된 깊이 영상 모델로부터 지식 증류를 통한 안정적인 3차원 휴먼 형상 복원 과정을 설명하기 위한 도면이다.

[0065] 본 발명의 실시예에 따르면 PIFu의 안정적인 휴먼 형상 복원을 위한 새로운 학습 방법을 제시한다. 최근 비변수 기반 3차원 휴먼 형상 복원 연구의 경우 다양한 공간 정보의 추론 과정을 학습 과정에 추가하여 안정적인 결과 출력을 구현한다. 이러한 방식은 PIFu의 기본 네트워크에 새로운 네트워크 구조를 추가하여 구현하므로 모델의 구동을 더 많은 수행 시간을 요구한다. 이에 발명에서는 OpenGL을 통해 휴먼 깊이 영상 생성 및 Hoffman의 깊이 영상 증류 방식을 이용하여 동일한 네트워크 구조로부터 안정적인 결과 출력이 가능한 깊이 영상 증류 기반의 학습 방식을 제안한다.

[0066] 도 6을 참조하면, 먼저 깊이 영상을 입력 받아 형상 추론을 구현하는 딥-음함수 네트워크를 학습한다(610). 딥-음함수 네트워크 학습은 기존 RGB 네트워크와 첫 번째 레이어를 제외하고 모두 동일한 구조를 가지지만, 깊이 영상에서 3차원 정보를 직접 입력 받으므로 매우 높은 성능의 복원 결과를 출력한 다. 본 발명에서는 효율적인 학습 성능 증가 확인을 위하여 전면 깊이 영상 입력 외에도 전, 후면 깊이 영상을 함께 입력 받는 네트워크에서의 학습을 진행하며 이에 대한 형상 복원 성능을 비교한다.

[0067] $$L_{hallucinate}(l) = \| \sigma(A_l^{depthNet}) - \sigma(A_l^{hNet}) \|_2^2 \quad (3)$$

[0068] 다음으로는, 사전 학습된 깊이 입력 모델로부터 RGB 영상을 입력으로 받는 환각 네트워크를 증류 학습한다(620). 두 네트워크는 영상을 입력 받는 첫 번째 레이어를 제외하고는 모두 동일한 구조를 가진다. 따라서 깊이 입력 모델로부터 환각 네트워크로의 지식 증류를 위해 식(3)과 같이 네트워크 간 ReLU, leaky ReLU, sigmoid의 모든 활성 함수 레이어에서의 손실함수를 계산하고 이를 학습에 활용하고자 한다. 활성 함수의 입력값에 sigmoid 값을 취하고 이를 통해 네트워크를 최적화한다. 이러한 학습 방식을 통해 본 발명에서는 동일한 네트워크에서의 안정적인 시스템 모델 구동을 구현한다. 각 점의 RGB 값을 출력하는 텍스쳐 추론 네트워크는 휴먼 형상 복원 네트워크와 유사하게 ResNet 인코더와 MLP가 합쳐진 구조이며, 인코더의 특징 맵은 형상 복원 인코더의 특징 맵과 합쳐져 MLP에 입력된다. 출력 결과는 3채널의 RGB 값을 추론하도록 학습된다. 실제 구동 시에는 깊이 영상 네트워크를 제외한 RGB 입력의 환각 네트워크만으로 0.5 레벨 집합 추론을 수행하며 텍스쳐 추론

네트워크, 후처리 알고리즘과 함께 3차원 휴먼 메쉬 형상을 복원한다.

[0069] 제안하는 깊이 영상 입력 네트워크는 PIFu와 동일하게 유클리디안 손실 함 수를 통해 학습한다. 3차원 메쉬 데이터의 주변에 무작위로 생성된 n개의 포인트 클라우드를 딥-음함수에 입력하여 0.5 레벨 집합 값의 평균 차이를 통해 모델을 학습한다. 제안하는 방식은 0 또는 1을 추론해야 하므로 유클리디안 손실 함수를 사용하여 딥-함수 f를 학습한다.

[0070]
$$L_{geometry} = \sum_{i=1}^{n} \| f_{gt}(P_i, I) - f(P_i, I) \|_2 \quad (4)$$

[0071] 환각 네트워크는 식(5)와 같이 앞서 활용한 유클리디안 손실 함수와 함께 깊이 영상 모델의 지식 증류를 위해 환각 손실함수를 합쳐 사용하여 네트워크를 학습한다.

[0072]
$$L = \lambda_1 * L_{geometry} + \lambda_2 * L_{hallucinate} \quad (5)$$

[0073]
$$L_{hallucinate} = \sum^{l} \| \sigma(A_l^{depthNet}) - \sigma(A_l^{RGBNet}) \|_2 \quad (6)$$

[0074] 환각 손실 함수는 Hoffman[2]이 사용한 손실 함수와 동일하게 사용한다. Hoffman은 사전 학습된 깊이 영상 기반의 모델과 학습할 RGB 기반 네트워크의 활성 함수 간 최적화를 통해 지식 증류를 구현한다. 각 네트워크는 손실 함수 이후의 특징 맵의 비교를 진행하여 이를 통한 모델 최적화를 구현한다. 이때 각 활성 함 수 이전 입력되는 값에 sigmoid 함수를 취한 결과 간의 손실 함수 계산을 통한 최적화를 진행한다. 본 발명에서는 해당 특징 맵의 유클리디안 차이를 통해 최적화를 진행한다. 해당 손실 함수는 네트워크 내 모든 활성 함수 이후에 동일하게 작동 한다. 이를 통해 깊이 영상 기반의 모델부터 RGB 입력 영상 모델로의 증류가 구현된다.

[0075]
$$L_{texture} = \sum_{i=1}^{n} \sum_{color}^{3} \| f_{gt}(P_i, I) - f(P_i, I) \|_1 \quad (7)$$

[0076] 텍스쳐 추론은 식(7)과 같으며 형상 출력과 동일하게 네트워크에 입력된 점 간의 차이를 구해 손실 함수를 계산하며, 이때는 L1 손실 함수를 사용하여 학습한다.

[0077] 이와 같이, 본 발명의 실시예들에 따르면 3차원 휴먼 형상 복원 방법 및 시스템을 통해 휴먼 깊이 영상 데이터셋 구축하고, 깊이 영상 기반의 딥-음함수 네트워크 학습한 후, 깊이 영상 네트워크를 활용한 RGB 입력 영상에서의 지식 증류를 수행하며, 비변수 기반 방식의 신경망 학습 과정에서 추가적인 공간 정보인 휴먼 깊이 영상을 이용하여 객체의 안정적인 형상을 실시간으로 복원할 수 있다.

[0079] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0080] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상

장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

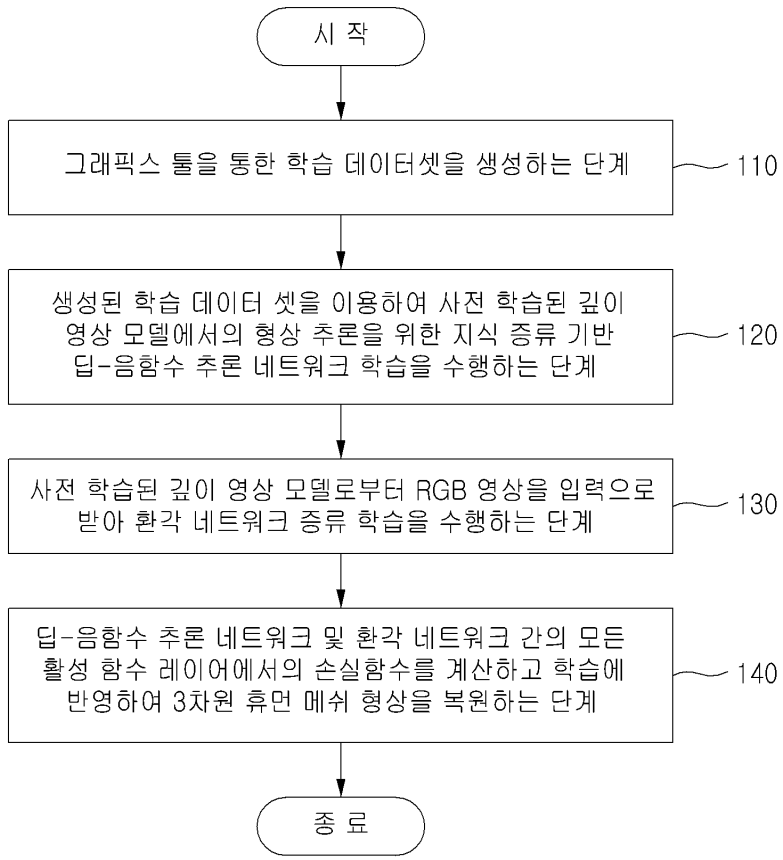
[0081] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0082] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

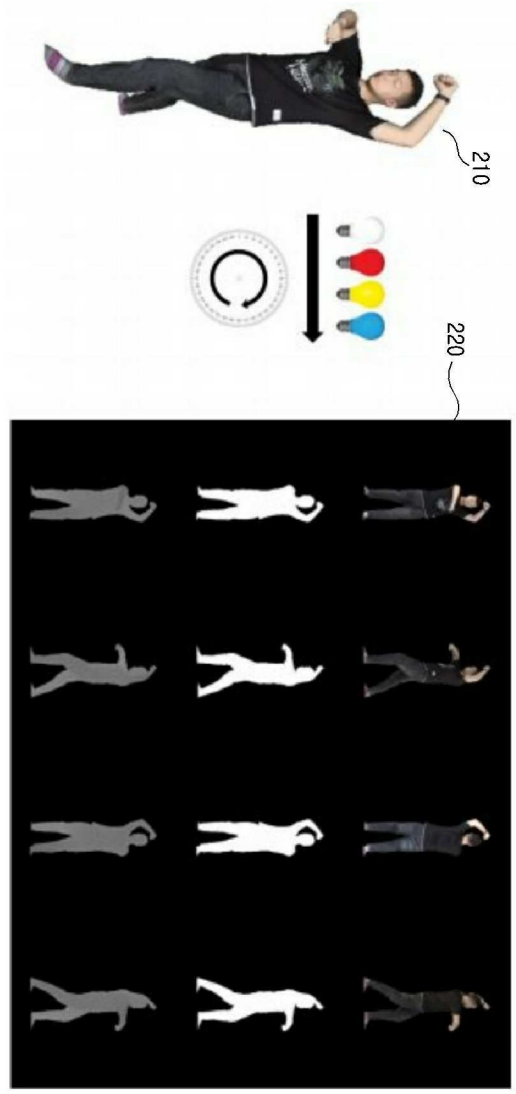
[0083] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

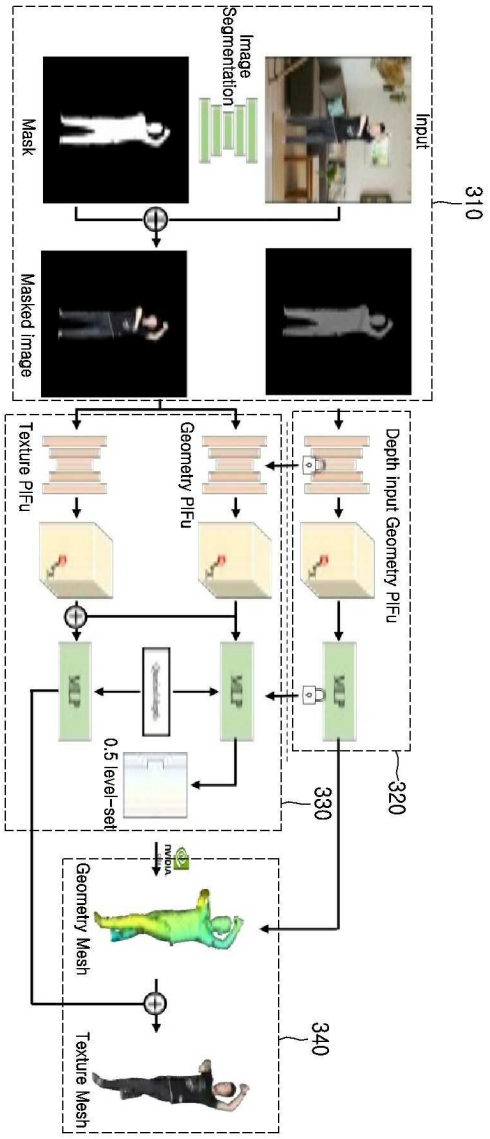
도면1



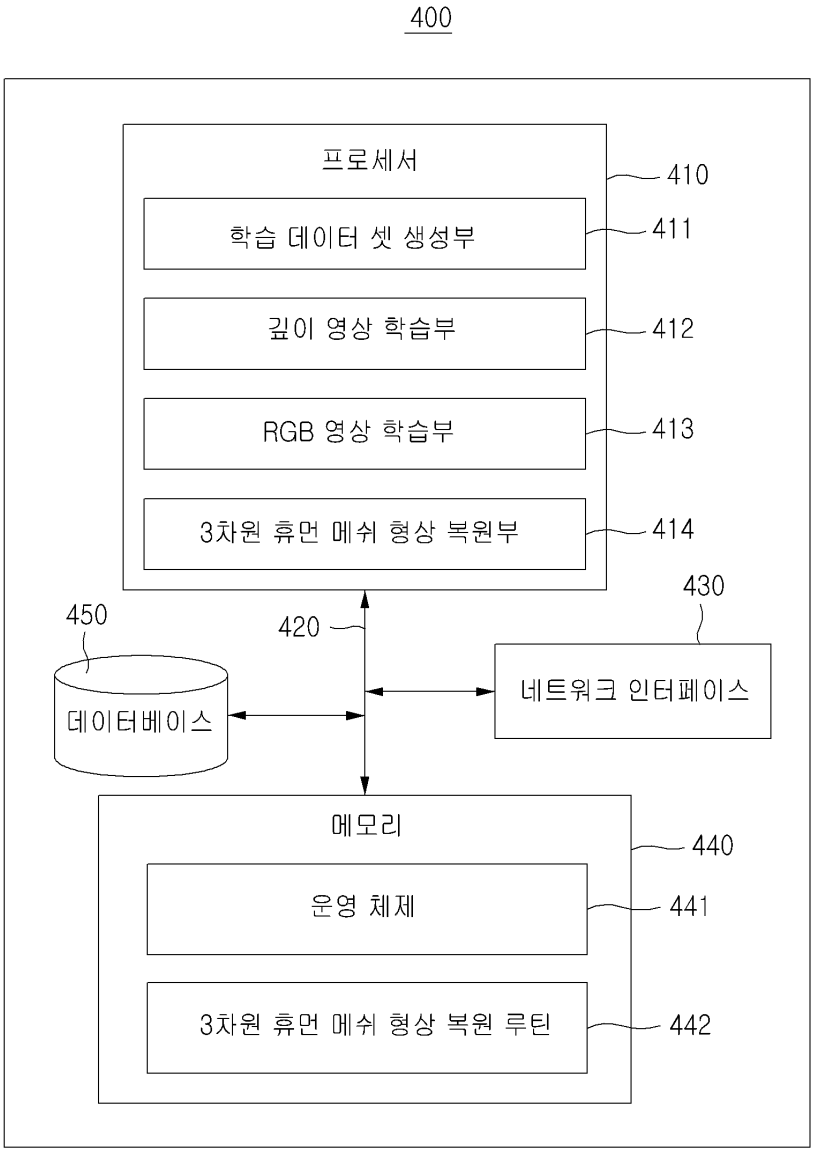
도면2



도면3


Image Segmentation
Mask
Input
Masked Image
310
Texture PIFu
Geometry PIFu
Depth Input Geometry PIFu
MLP
MLP
MLP
0.5 level-set
330
320
Geometry Mesh
Texture Mesh
340

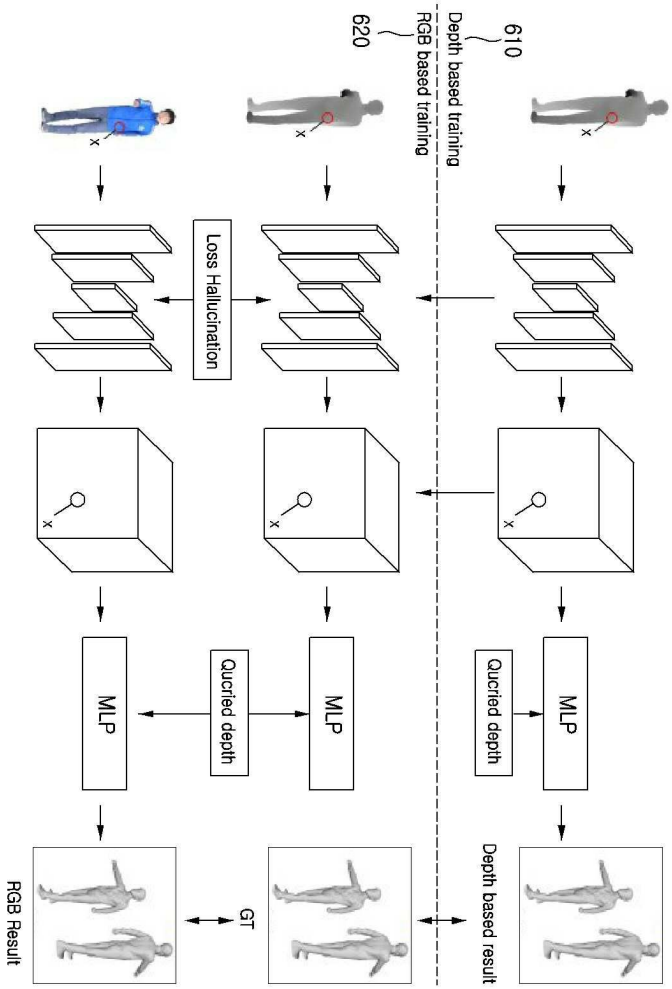
도면4


400
프로세서
410
학습 데이터 셋 생성부
411
깊이 영상 학습부
412
RGB 영상 학습부
413
3차원 휴먼 메쉬 형상 복원부
414
430
450
420
네트워크 인터페이스
데이터베이스
메모리
440
운영 체제
441
3차원 휴먼 메쉬 형상 복원 루틴
442

도면5


Input video
(a)
Detection result
(b)
Segmentation result
(c)

도면6


620
610
RGB based training
Depth based training
Loss Hallucination
MLP
Qucried depth
MLP
Qucried depth
MLP
RGB Result
GT
Depth based result