딥러닝을 활용한 다중 시점 스트레오와 3차원 자세 추정을 결합한 사람 표면 복원에 대한 연구

이혁상, 김재경, 이상훈

연세대학교

melungl@yonsei.ac.kr, jkkproject@yonsei.ac.kr, slee@yonsei.ac.kr

A Study on human surface reconstruction by integrating deep learning with multi-view stereo and 3D pose estimation

Lee Hyucksang, Kim Jaekyung, Lee Sanghoon

Yonsei Univ.

요 약

딥러닝을 활용한 3차원 사람 복원 기술은 영화, 게임, 메타버스 등 다양한 분야에서 주목받고 있는 기술이다. 본 연구는 딥러닝을 기반으로 한 다중 시점 스테레오를 활용하여 사람의 표면을 더 정확하고 효율적으로 복원하는 방법을 탐구한다. 기존의 다중 시점 스트레오를 통해 사람을 복원하기 위해서는 사람이 직접 깊이에 대한 범위를 일일이 조절해야 하여 많은 시간 투자가 필요하다는 단점이 존재한다. 하지만 본 연구에서는 이를 3차원 자세 추정 알고리즘과 결합하여 3차원 관절 위치를 깊이 범위에 대한 초기 정보로 활용함으로써 사람의 적합한 깊이 범위를 보다 빠르게 찾아, 사람의 정밀한 깊이를 추정하고, 이를 통해 효율적으로 사람의 정밀한 3차원 표면을 복원하는 방법을 제시한다.

Ⅰ. 서 론

현대 컴퓨터 비전 기술을 활용한 현실 세계의 객체를 가상 세계로 이동시키는 3차원 재구성 기술은 영화, 게임, 메타버스와 같은 다양한 분야에서 큰 관심을 받고 있다. 특히, 현실 세계의 사람을 가상 세계로 옮기는 기술은 해당 분야에서 핵심 기술로 그 활용 가능성은 무궁무진하다. 이러한 사람을 디지털 정보로 복원하는 기술은 딥러닝 기술의 발전으로 인해 더욱 정확하고 신속하게 사람을 복원할 수 있게 되었다.

본 논문에서는 3차원 객체 표면을 복원하기 위한 방식 중에서 Plane Sweep[1] 방식의 딥러닝 기반 다중 시점 스트레오[2]를 활용한다. Plane Sweep은 다중 시점 이미지를 사용하여 각 이미지의 깊이를 추정하는 방법으로, 이미지를 3차원 공간 상에 투영하여 깊이를 추정하는 방법을 사용한다. 이를 통해 기존의 2차원 상의 특징점 정합 방법인 베이스라인 스트레오[3] 방법보다 빠른 깊이 추정이 가능하다. 그러나 이러한 방식은 제한된 메모리 용량으로 인해 추정할 깊이 범위를 초기에 임의로 설정해 주어야 하는 단점이 있다. 이로 인해 사용자는 각 이미지의 복원한 대상에 대한 적절한 깊이 범위를 수동으로 설정하고 최적의 결과를 찾기 위해 번거로운 반복 작업을 수행해야 한다.

이미지에 나타난 사람의 유효한 깊이 범위의 경우 사람의 머리, 손, 발과 같이 말단의 3차원 위치 정보가 강력한 도움을 줄 수 있다. 이에 따라 본 논문에서는 각 이미지에서 사람에 대한 대략적인 깊이 범위를 알기 위해 사람의 3차원 자세에 주목하였다. 최근의 사람 자세 추정 알고리즘은 딥러닝의 발전으로 인해 빠르고 정확하게 추정할 수 있게 되었다. 따라서 본 논문에서는 딥러닝 기반 사람 자세 추정 알고리즘[4]을 활용하여 여러 시점에서의 사람의 2차원 자세를 추정하고 이를 3차원 자세로 복원하여 얻은 정보를 각 이미지에 투영하여 명확한 초기 깊이 범위를 자동으로 설정하는 방식을 제안한다. 이를 통해 기존의 번거로운 설정 작업을 크게 간소화할 뿐만 아니라 기존 방법에 비해 더욱 정확한 3차원 사람 표면 복원을 수행한다. 더불어, 본 논문에서는 기존에 비효율적으로 진행되던 직렬 과정의 복원을 병렬로 구성하여 전반적인 시간 효율성을 향상시키는 방식을 제안하여 보다 빠른 복원 과정을 구현했다.

Ⅱ. 본론

본 논문에서는 3차원 사람 표면을 복원하기 위해 딥러닝 기반 다중 시점 스트레오[2]를 통해 각 시점 이미지의 사람에 대한 깊이 값을 추정한다. 이 때 정밀한 깊이 추정을 위해서는 각 이미지에서 가장 유효한 초기 깊이 범위값을 지정해 주어야 한다.

이를 위해 본 논문에서는 3차원 자세 정보를 이용하고자 하고, 먼저 이미지상에서 사람의 2차원 자세를 추정하는 Google사의 딥러닝 자세 추정 모델인 mediapipe[4]를 활용하여 다중 시점 이미지에서의 사람의 33개 관절에 대한 2차원 자세를 추정한다. 다중 시점에서 사람의 자세를 추정할 경우, 사람의 앞면 추정 정확도는 높으나 측면과 후면의 자세 추정의 정확도가 급격히 떨어지는 문제가 발생한다. 만약 이러한 정확도가 낮은 2차원 자세 정보를 기반으로 3차원 자세를 재구성하면, 부정확한 3차원 자세를 복원하게 된다. 따라서 본 논문에서는 관절의 추정 신뢰도를 함께 사용하여, 일정 신뢰도 이상의 관절의 2차원 위치 정보와 캘리브레이션된 카메라 값을 활용하여 그림1의 (가)와 같이 정확한 3차원 자세를 추정할 수 있도록 하였다.

추정된 3차원 자세를 각 시점의 이미지에 투영하면 이미지에서의 각 관절이 가지는 깊이 값을 알 수 있다. 본 논문에서는 이미지에서의 33개 관절에 대한 깊이 값을 비교하여 가장 가까운 깊이 값과 가장 먼 깊이 값의 사이 범위를 사람에 대한 유효 범위로 지정하여 해당 정보를 딥러닝 기반 다중 시점 스트레오의 입력으로 넣어주었고, 그 결과 그림1의 (나)와 같이

각 시점 이미지의 정밀한 깊이 정보를 추정할 수 있었다.

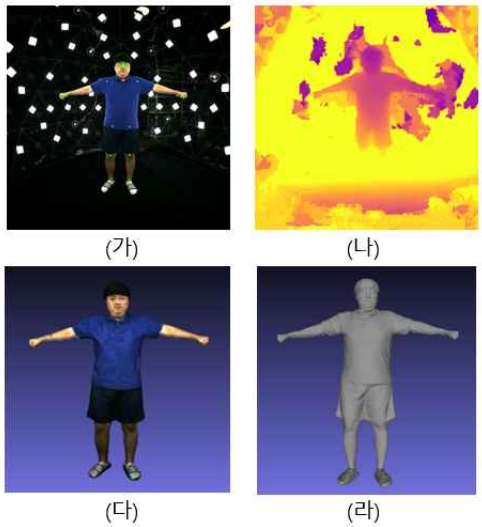

그림 1. (가) 3차원 자세 추정 결과, (나) 다중 시점 스트레오 깊이 결과, (다) 복원된 3차원 사람 포인트클라우드, (라) 복원된 3차원 사람 표면

제안한 자세기반 깊이 초기 정보가 기존의 방법에 비해 얼마나 성능 향상에 기여했는지 비교 분석하기 위해 정성 및 정량 평가를 진행하였다. 각 방법을 통해 얻은 이미지의 깊이 정보를 활용하여 3차원 공간 상에 정합하는 과정을 거쳐 3차원 점들인 포인트클라우드로 3명의 사람에 대해 시각화하였다. 먼저 그림2의 왼쪽 결과와 같이, 기존 방법을 통해 복원된 3차원 사람의 경우 얼굴과 손 등 사람의 말단 영역의 3차원 복원이 미흡한 결과를 보여주었다. 하지만 제안한 자세 추정기반으로 깊이 범위를 적절하게 주어진 결과의 경우 이러한 영역에서 강인한 성능 향상을 보여주었다. 정량 평가 결과 또한 표 1과 같이, 기존 방법은 평균 2,7,91,723개의 3차원 점들이 재구성되었지만, 제안한 방법의 경우 4,144,958개의 3차원 점들이 재구성되어 기존 방법에 비해 제안한 방법이 1.48배 정도 더 밀집한 포인트클라우드 결과를 보여주었다.

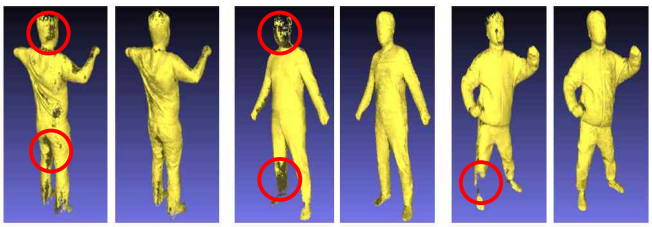

그림 2.. 기존 방식의 3차원 사람 복원 결과 (좌) 자세추정 기반으로 초기 깊이 값을 지정해준 3차원 사람 복원 결과 (우) 비교

Pointcloud 수	기존	자세추정 기반
Subject 1	2,216,283	2,725,945
Subject 2	3,112,237	4,911,449
Subject 3	3,046,650	4,797,482
평균	2,791,723	4,144,958

표 1. 자세추정 기반으로 초기 깊이 값을 지정하여 3차원 사람 복원 결과표

다중 시점 스트레오를 통해 3차원 사람 표면을 복원하는 과정은 크게 3가지로 나눌 수 있다. 첫 번째는 다중 시점 이미지로부터 깊이 정보를 추정하는 과정, 두 번째는 그림 1의 (다)와 같이 각 시점의 깊이 정보를 3차원의 점들인 포인트 클라우드로 정합하는 과정, 마지막으로 그림1의 (라)와 같이 정합된 3차원 점들로부터 표면을 복원하기 위한 Poisson Surface Reconstruction[5] 과정이 있다. 그러나 기존의 딥러닝 기반 다중 시점 스트레오[2]는 각 과정이 직렬로 연결되어 여러 프레임의 이미지, 즉 영상에 대한 사람을 복원할 때 시간적으로 비효율적이라는 단점이 존재한다. 따라서 본 논문에서는 각 과정에 대해 병렬로 재구성하여 각 단계가 병행적으로 실행될 수 있도록 설계하였다. 이러한 과정을 통해 아래 표 2와 같이 기존의 과정에 비해 2배 이상 빠른 실행속도를 달성할 수 있었다.

	기존	병렬 프로세스
프레임 당 평균 실행시간	25 sec	12 sec

표 2. 병렬 프로세스 기반 Frame당 3차원 사람 복원 평균 실행 시간

Ⅲ. 결론

본 논문에서는 다중 시점 이미지 기반의 딥러닝 3차원 표면 복원 애플리케이션[2]을 활용하여 사람의 표면을 효과적으로 복원하기 위한 새로운 방법을 제안하였다. 이를 위해 딥러닝 기반의 사람 자세 추정 모델[4]을 활용하여, 사람의 3차원 관절 정보를 추정하고, 이 정보를 각 이미지마다 초기 깊이 값으로 설정하여 사람의 3차원 표면 복원을 보다 정확하고 효율적으로 수행하였다. 제안한 방법은 기존 방법이 제대로 처리하지 못한 머리, 손, 발 영역을 보다 효과적으로 복원할 뿐만 아니라 기존 방법에 비해 약 1.48배 밀집한 사람 표면의 포인트클라우드를 복원할 수 있었다. 또한 각 과정을 병렬로 재구성하여 기존 방법에 비해 2배 이상 빠른 속도로 영상에 대한 3차원 사람 표면을 복원할 수 있었다.

ACKNOWLEDGMENT

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1A2C3011697).

참 고 문 헌

[1] Collins, Robert T. "A space-sweep approach to true multi-image matching." Proceedings CVPR IEEE computer society conference on computer vision and pattern recognition. Ieee, 1996.

[2] Wang, Xiaofeng, et al. "MVSTER: Epipolar transformer for efficient multi-view stereo." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[3] Bleyer, Michael, Christoph Rhemann, and Carsten Rother. "Patchmatch stereo-stereo matching with slanted support windows." Bmvc. Vol. 11. 2011.

[4] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).

[5] Kazhdan, Michael, Matthew Bolitho, and Hugues Hoppe. "Poisson surface reconstruction." Proceedings of the fourth Eurographics symposium on Geometry processing. Vol. 7. 2006.