단일 LiDAR를 활용한 End-to-End 기반 3D 모델 생성 방법

곽정훈, 성연식*
동국대학교 멀티미디어공학과
{jeonghoon, sung}@dongguk.edu

End-to-End based 3D Model Generation Method using a Single LiDAR

Jeonghoon Kwak, Yunsick Sung*
Dept. of Multimedia Engineering, Dongguk University-Seoul

요 약

원격 및 가상환경에서 사용자의 동작에 따른 3D 모델을 제공하기 위해 light detection and range (LiDAR)로 측정된 3D point cloud로 사용자의 3D 모델이 생성되어 원격 및 가상환경에 사용자의 모습이 제공된다. 하지만 3D 모델을 생성하기 위해서는 사용자의 신체 전부가 측정된 3D point cloud가 필요하다. 사용자의 신체 전체를 측정하기 위해서는 적어도 두 개 이상의 LiDAR가 필요하다. 두 개 이상의 LiDAR을 사용할 경우에는 LiDAR을 사용할 공간과 LiDAR를 구비하기 위한 비용이 발생한다. 단일 LiDAR로 3D 모델을 생성하는 방법이 요구된다. 본 논문에서는 단일 LiDAR에서 측정된 3D point cloud를 이용하여 3D 모델을 생성하는 방법이 제안된다. End-to-End 기반 Convolutional Neural Network (CNN) 모델로 측정된 3D point cloud를 분석하여 사용자의 체형과 자세를 예측하도록 학습한다. 기본자세를 취하는 동안 수집된 3D point cloud로 기본이 되는 사용자의 3D 모델을 생성한다. 학습된 CNN 모델을 통하여 측정된 3D point cloud로 사용자의 자세를 예측하여 기본이 되는 3D 모델을 수정하여 3D 모델을 제공한다.

1. 서론

사용자의 모습을 가상 및 원격환경에서 활용하기 위하여 사용자의 모습을 제공하기 위한 3D 모델이 생성된다. 3D 모델을 생성하기 위하여 가상 및 원격환경에서 사용할 사용자의 3D 모델을 Light Detection And Range (LiDAR) 또는 깊이 카메라로 측정된 3D point cloud로 사용자의 3D 모델이 생성되어 제공된다[1]. 측정된 3D point cloud로 3D 모델의 외형이 결정된다. 3D point cloud로 3D 모델을 구성하기 위해서는 최소 2개 이상의 센서로 3D point cloud의 측정이 필요하며, 사용자의 신체의 360도를 측정하기 위한 공간도 요구된다. 다중 센서를 사용하여 3D 모델을 생성하는 과정에서 따라 발생하는 비용 또는 공간적인 문제를 해결하기 위하여 단일 센서로 실시간으로 사용자에 동작에 따른 3D 모델을 생성하는 방법이 필요하다.

단일 센서로 측정된 3D point cloud과 사전에 사용자 동작에 따라 3D 모델을 제공하는 템플릿을 사용하여 3D 모델을 생성한다[2]. 측정된 3D point cloud가 템플릿에 포함되는 3D 모델 중에서 가장 유사한 3D 모델을 찾아 3D 모델을 제공한다. 그러나 3D 모델을 생성하는 과정에서 다양한 사용자의 신체 또는 사용자의 다양한 동작에 사전에 정의하기에는 한계가 발생한다. 사용자의 다양한 동작에 3D 모델 템플릿을 사전에 구축하지 않고 3D 모델을 제공하는 방법이 필요하다.

단일 카메라로부터 촬영된 단일 이미지와 3D 모델 템플릿을 활용하여 사용자의 동작에 맞게 3D 모델이 제공된다[3]. 단일 이미지로부터 사용자의 자세와 체형을 예측하고 예측된 자세 및 체형에 따라 3D 모델이 보정된다. 하지만 단일 이미지로부터 사용자의 자세를 추정하기 위한 방법이 필요하다. 사용자의 자세를 사전에 지정하지 않고 사용자에 자세

* 교신저자: 성연식 (sung@dongguk.edu)

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음"(2019-0-01585)

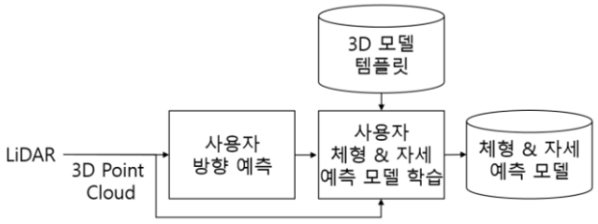
에 맞게 3D 모델을 제공하는 방법이 요구된다.

본 논문은 3D 모델 템플릿을 활용하여 단일 3D point cloud를 기반으로 3D 모델을 생성하는 방법을 제안한다. 3D point cloud가 End-to-End 기반 Convolutional Neural Network (CNN) 모델에 입력되어 3D 모델의 체형과 자세를 예측할 수 있도록 학습한다. CNN 모델은 3D 모델이 3D point cloud와 일치할 수 있게 학습된다. 실행 시에는 사용자의 외형과 유사한 기본 3D 모델을 생성한다. 학습된 CNN 모델을 측정된 3D point cloud를 활용하여 사용자의 자세를 예측한다. 예측된 자세를 기반으로 기본 3D 모델을 보정함으로써 사용자의 3D 모델이 제공된다.

2. 3D Point Cloud 기반 3D 모델 생성 방법

사용자로부터 측정된 3D point cloud를 이용하여 사용자의 체형과 자세에 맞는 3D 모델을 생성하기 위해 CNN 모델을 학습하는 과정은 그림 1과 같다. 사용자 방향 예측에서는 사용자의 신체가 전방이 되는 방향을 예측한다. 3D 모델은 전체를 제공되지만 3D point cloud는 사용자의 신체 일부가 측정되기 때문에 사용자 신체에 맞게 회전이 필요하다. 예측된 사용자 방향을 활용하여 사용자 신체의 전방과 같이 3D 모델 템플릿의 3D 모델의 방향이 동일하게 3D 모델의 방향을 변경한다.

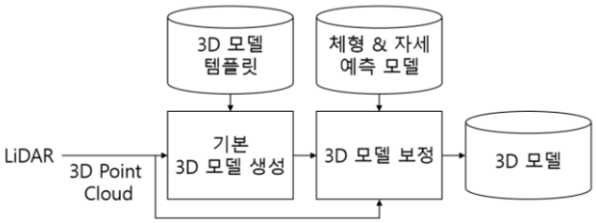
사용자 체형 & 자세 예측 모델 학습에서는 3D 모델 템플릿을 활용하여 CNN 모델을 학습한다. 3D 모델 템플릿은 SMPL 모델[4]과 같이 다양한 남성 및 여성의 체형이 3D 모델을 제공되며 스켈레톤을 기반으로 체형과 자세를 변형할 수 있다. CNN 모델은 입력된 사용자의 3D point cloud를 이용하여 3D 모델의 체형과 자세를 예측할 수 있게 학습된다. 3D point cloud에 포함된 위치들과 3D 모델간의 유클리디안 거리로 차이가 계산된다. 계산된 차이를 이용하여 CNN 모델이 수정된다.



(그림 1) 체형과 자세를 예측하기 위한 학습과정

사용자로부터 측정된 3D point cloud를 이용하여 사용자의 체형과 자세에 따른 3D 모델을 생성하는 과정은 그림 2과 같다. 기본 3D 모델 생성에서는 사용자의 기본 3D 모델을 생성한다. 기본자세를 수행하는 과정동안 전체 신체의 3D point cloud를 수집한다. 수집된 3D point cloud과 유사한 3D 모델 템플릿을 찾고 사용자의 신체 특성에 맞게 보정된다.

3D 모델 보정에서는 학습된 CNN 모델을 활용하여 3D point cloud를 이용하여 사용자의 자세가 예측된다. CNN 모델에 3D point cloud의 체형과 자세가 예측된다. 하지만 체형 정보는 이미 사용자의 신체에 맞게 추론하였기 때문에 사용하지 않고 자세의 값만 활용한다. 기본 3D 모델을 현재 사용자의 자세에 맞게 수정하여 사용자의 모습을 제공하는 3D 모델이 제공된다.



(그림 2) 체형과 자세를 제공하는 3D 모델 생성과정

사사표기

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음"(2019-0-01585)

참고문헌

[1] Kim, Y., Baek, S., Bae, B., “Motion Capture of the Human Body Using Multiple Depth Sensors,” ETRI Journal. Vol. 39, No. 2, pp. 181-190, 2017.

[2] Zhao, T., Li, S., Ngan, K.N., Wu, F., “3-D Reconstruction of Human Body Shape from a Single Commodity Depth Camera,” IEEE Transactions on Multimedia, Vol. 21, No. 1, pp. 114-123, 2019.

[3] Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J., “End-to-end Recovery of Human Shape and Pose,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018, pp. 7122-7131.

[4] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black. and M.J., “SMPL: A Skinned Multi-person Linear Mode,” ACM Transactions on. Graphics, Vol. 34, No. 6, pp.1-16, 2015.