춤추는 아바타: 당신도 싸이처럼 춤을 출 수 있다.

구동준, 주영돈, 브이 반 만, 이정우, *안희준[1]
서울과학기술대학교
*heejune@seoultech.ac.kr

Dancing Avatar: You can dance like PSY too.

Dongjun Gu, Youngdon Joo, Vu Van Manh, Jungwoo Lee, Heejune Ahn
Seoul National University of Science and Technology

요 약

본 논문에서는 사람을 키넥트로 촬영하여 3 차원 아바타로 복원하여 연예인처럼 춤을 추게 하는 기술을 설계 구현하였다. 기존의 순수 딥러닝 기반 방식과 달리 본 기술은 3 차원 인체 모델을 사용하여 안정적이고 자유로운 결과를 얻을 수 있다. 우선 인체 모델의 기하학적 정보는 3 차원 조인트를 사용하여 추정하고 DensePose 를 통하여 정교한 텍스쳐를 복원한다. 여기에 3 차원 포인트-클라우드와 ICP 매칭 기법을 사용하여 의상 모델 정보를 복원한다. 이렇게 확보한 신체 모델과 의상 모델을 사용한 아바타는 신체 모델의 rigged 특성을 그대로 유지함으로써 애니메이션에 적합하여 PSY 의 <강남스타일>과 같은 춤을 자연스럽게 표현하였다. 개선할 점으로 인체와 의류 부분의 좀 더 정확한 분할과 분할과정에서 발생할 수 있는 노이즈의 제거 등을 확인되었다.

1. 서론

최근 4 차산업혁명의 영향과 더불어 코로나 시대가 된 지금 비대면 활동은 굉장히 중요한 요소가 되었다[1]. 그렇기에 우리는 단순히 문자와 음성으로 소통하는 것을 뛰어넘어 비대면에서도 실제로 사회생활을 하는 것과 같이 나를 닮은 아바타가 움직이고 상호작용을 하며 여가시간을 즐길 수 있도록 인간의 모습을 모방한 아바타를 생성하고 그것이 자연스럽게 움직일 수 있게 하는 기술을 연구했다.

아바타를 만드는 연구는 다양한 방법으로 시도 되고 있다. 대부분은 딥러닝을 사용한 포즈전달기술의 응용[2]으로 연구되고 있다. 그러나 딥러닝 기술의 완성도가 아직 높지 않아 수준의 화질이나 자세가 제어가 안정적이지 않거나 화질이 떨어지는 모습을 보이는 경우가 많다. 반면 모델을 기반으로 하는 경우 [3] 는 모델의 복원에 상당한 수작업이 필요하여 자동화에 어려운 상황이다.

본 연구는 모델을 기반으로 하는 방식을 취하되 이때 발생하는 기술적인 문제인 이미지 및 포인트 클라우드와 3 차원 모델간의 매칭의 문제를 해결하는 것을 목표로 하였고 이를 바탕으로 무보 (춤 악보)를 사용하면 어떤 대상이던 원하는 춤을 출 수 있는 시스템을 설계하고 구현하였다. 이를 위하여 SMPL 3 차원 인체모델[4]을 모델을 사용하고 Densepose[5]를 통한 매칭 알고리즘과 최종적으로 ICP 기법을 이용한 의상을 착용한 모델을 복원하는 단계들을 설계하고 구현하였다.

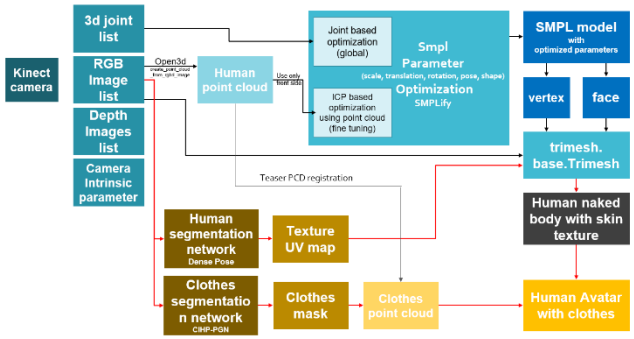
본 논문의 구성은 다음과 같다. 2 절에서는 시스템의 전체적인 구성과 키넥트 카메라를 사용한 데이터 획득 방법, 3 차원 조인트를 사용한 개선된 모델 파라메터 SMPLify 기술[6], 텍스쳐를 입힐 수 있게 정교하게 텍스쳐와 매핑해 주는 Densepose 기술, 인체 모델과 옷 모델이 합쳐진 최종 모델과 그것을 애니메이션으로 만들 기술에 대해 소개한다. 3 절에서는 실험 이미지를 통해 단계별 결과 및 최종 모델, 그리고 모델이 움직이는 애니메이션 영상을 보인다. 4 절에서는 성능평가를 실시하고 발견된 문제점들을 분석해 개선할 수 있는 방법에 대해 논의한다.

2. 제안 방법

Fig.1 은 전체 시스템의 구성으로 키넥트 카메라로 3 차원 조인트 정보, RGB 이미지 정보, 깊이이미지 정보, 카메라 모델

[1] 캡스톤 지도교수

을 얻는다. 우선 의상을 입지 않은 인체 모델을 얻는 단계이다. 3 차원 조인트 정보를 이용해 SMPL 모델을 일차 추정을 한다. 정교한 정합을 위하여 ICP 를 이용해 SMPL 모델을 최적화하면 사람의 체형, 자세 등을 포함한 사람의 신체의 기하 모델이 생성되게 된다. 그 모델에 Densepose 를 이용해 RGB 이미지 상의 신체를 부위별로 나눈 후 SMPL 모델의 점과 면에 매핑해 텍스쳐의 위치를 지정해 텍스쳐를 확보한다. 여기에 의상을 모델링하기 위하여 RGB 이미지에서 CIHP-PGN[7] 이미지 분할 네트워크를 이용해 의상영역 분리 및 모델을 생성하여 의상을 갖춘 3 차원 아바타를 생성하게 된다. 이렇게 얻은 모델은 자세 값만 정해주면 어떠한 율동도 가능하다.



<Fig.1 전체 시스템 구성>

2.1 인체 모델 생성

SMPL 인체 모델[4]은 자세와 체형에 대한 파라메터를 통해 다양한 인체를 표현할 수 있다. 따라서 입력데이터에서 SMPL 모델의 파라메터를 추정하는 단계가 필요하다. 키넥트 카메라에서 얻은 3 차원 조인트 정보를 이용해 1 차적으로 SMPL 모델의 파라메터를 추정한다. 이 때 조인트 정보를 통한 최적화만으론 양질의 모델을 얻을 수 없어 ICP 기술을 통한 최적화를 추가로 이용했다. 2 차 최적화를 위해 사용할 ICP 기술은 강체에 적용하는 경우 SVD 를 사용한 이동 변환 행렬을 구하는 대신 비선형 최적화를 사용한 SMPL 모델과 카메라의 파라메터를 계산하도록 변형하였다. 이 단계에서는 전면부에서 키넥트 촬영한 한 장만을 사용하였다.

2.2 인체 및 의상 텍스쳐 추출

앞서 얻은 SMPL 모델은 기하정보만을 가지고 있다. 3 차원 모델의 텍스쳐를 확보하기위해서는 SMPL 모델과 RGB 이미지간의 매핑정보가 필요하다. Denspose [5]를 이용하면 RGB 이미지 픽셀을 SMPL 상의 UV 좌표계로 재구성된다. 즉, SMPL 모델의 vertex 또는 face 와 매핑을 구할 수 있다. 이때 편의상 전면부 이미지를 후면부에도 사용하여 전체 UV 맵을 확보하였다.

2.3 의상 모델 생성

앞 선 단계를 통하여 얻은 결과는 대상의 의상을 포함하는 모델이 아니라 신체에 대한 모델로 렌더링 결과는 바디페인팅을 한 형태로 나타나게 된다. 따라서 의상 모델을 생성할 필요가 있었다. 키넥트 카메라에서 얻은 여러 장의 RGBD 이미지를 통해 3 차원의 포인트 클라우드를 생성하였다. 먼저 RGBD 이미지에서 전경의 옷 부분을 추출하기 위해 CIHP-PGN 기술을 이용하였다. 여러장의 포인트 클라우를 합치기 위하여 Teaser [8]를 통한 레지스트레이션을 이용하였다.

이 의상 영역에 해당하는 부분을 촬영된 3D 포인트 클라우드와 버틱스 대응점을 찾고 해당 대응점 간의 변위 값을 구하여 이를 SMPL 모델과 의상 모델간의 차이로 사용한다. 텍스쳐 정보는 포인트 클라우드의 색상 정보를 이용한다. 이렇게 만들어진 의상 메시와 앞서 구한 인체 메시를 합쳐서 휴먼 아바타를 구성한다.

2.4 아바타 애니매이션

3 차원 아바타의 애니매이션은 SMPL 모델이 rigged 모델이므로 자세정보를 변형하여 쉽게 얻을 수 있다. 또한 앞서 얻어진 의상 메시도 SMPL 에 기반하고 있으므로 마찬가지 방식으로 변형하고 변위를 추가해주는 방식으로 얻을 수 있다. 움직임 데이터는 BVH 파일을 이용하여 SMPL 모델과 매칭을 통하여 얻을 수 있다. 다양한 BVH 파일을 사용하였으나, 본 논문에서는 흥미를 위하여 PSY 의 <강남 스타일>을 사용한 결과를 보였다.

3. 실험 결과

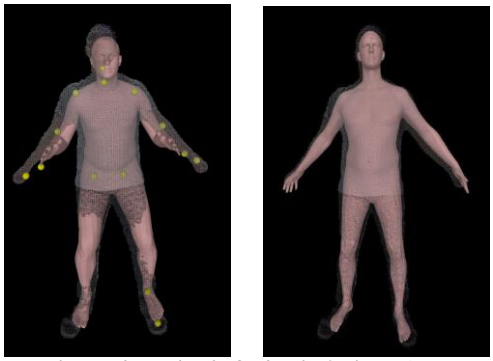
실험에 사용한 데이터는 본 연구의 참여자들을 촬영하여 얻었다. 촬영 시 팔을 벌리고 서 있는 A 자형 자세로 촬영하였고, 사진을 찍는 위치 또는 배경만 바꿔 보는 형태로 실험을 진행했다.

3.1 인체 SMPL 모델 생성 결과

먼저 SMPL 모델의 추정 1 단계에서 3 차원 조인트 정보를 통해 예측하므로 키넥트 3 차원 조인트 정보를 검증하였다. SMPLify 에 적용되었을 때 인체의 자세를 잘 잡아낼 수 있는지에 대하여 실험했는데 키넥트에서 나오는 조인트의 정보도 상당히 정확한 것으로 관측되었다. Fig. 2 는 입력 사진과 그 사진으로 만들어진 SMPL 모델을 합성한 사진이다.

<Fig.2 3D Joint 정보를 통해 생성한 SMPL 모델>

다음 단계는 A-포즈로 촬영한 사진과 SMPLify 를 통해 만들어지는 모델이 얼마나 비슷한 체형과 자세를 가지고 있는지 실험을 진행했다. 고정된 자세이므로 조인트를 통한 자세의 최적화는 잘 진행되었고, ICP 를 통한 최적화에서도 SMPL 모델이 인체의 실루엣에 근접하게 잘 늘어나 있는 것을 관찰할 수 있었다. Fig.3 은 조인트 정보를 통해 최적화를 하는 과정의 사진과 ICP 를 이용해 최적화를 마친 사진이다.



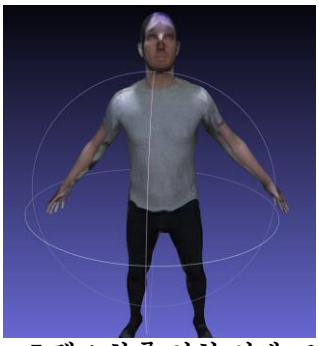
<Fig.3 (좌)조인트를 이용한 최적화, (우) ICP 를 이용한 최적화>

3.2 신체와 의상 텍스쳐링 결과

Fig. 4 는 실험에 사용된 RGB 이미지와 Desnepose 를 이용해 분할된 결과의 사진이다. RGB 이미지에서 Densepose 를 사용했을 때 신체가 각 부위별로 잘 나누어 졌는지 실험해 보았다. 실험은 연구실 복도에서 촬영한 사진으로 진행되었으며 실험 결과 머리와 팔의 상부 및 하부, 손, 몸통, 허벅지, 종아리, 발 등으로 신체부위의 분할이 잘 이루어진 것을 관찰할 수 있었다.

<Fig.4 (좌)RGB 이미지, (우)신체 분할 결과>

Fig. 5 는 Densepose UV 좌표를 사용하여 SMPL 모델에 텍스쳐를 입힌 결과이다. 피부가 아니라 옷의 텍스쳐가 입혀진 부분은 다음 단계에서 의상 메시를 만들어 옷을 입히면 가려져 보이지 않게 된다. 실험 결과 텍스쳐가 바른 위치에 잘 매핑되 칠해진 것을 관찰할 수 있었다. 팔의 밑부분에서 잘 못된 색이 칠해진 부분을 발견할 수 있었는데, 이 부분은 주변 조명 등의 영향으로 인한 그림자로 인해 생긴 노이즈인 것으로 추정되며 추후에 개선이 필요한 사항이다.



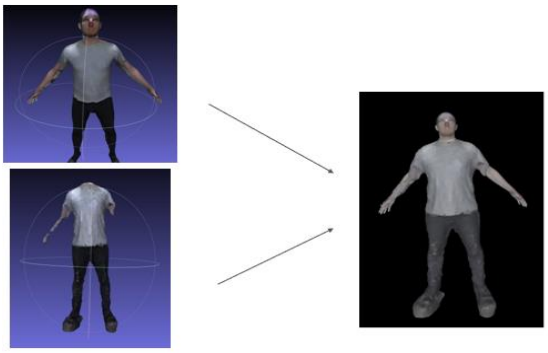
<Fig.5 텍스쳐를 입힌 인체 모델>

3.3 Clothes Mesh 와 최종 모델 생성 결과

Fig.6 은 CIHP-PGN 을 통해 의상영역을 분리한 결과와 그것을 통해 만들어진 의상 메시 사진이다. CIHP-PGN 사람 분할 네트워크를 이용해 RGB 이미지에서 옷부분을 추출한 결과 옷 부분을 잘 분리해내는 것을 관찰할 수 있었으며 의상 매시도 옷의 형태를 나타내는 것을 관찰할 수 있었다. 단, 의상메시의 팔 부분에서 노이즈가 관찰되었는데 이 노이즈 또한 그림자에 의한 것으로 예상된다. 옷의 색과 그림자의 색이 비슷해 생기는 것으로 추정되며 추후 노이즈 제거 등을 통해 개선해야 될 사항이다.

<Fig.6 (좌)옷 부분 분리 결과, (우)생성된 Clothes Mesh>

Fig. 7 은 최종적으로 만들어진 Human Avatar 의 사진이다. 위에서 생성한 의상 메시를 앞서 만든 텍스쳐가 입혀진 인체 모델에 입혀 옷을 입은 형태의 완성된 Human Avatar 를 만들었다. 실험 결과 사람이 옷을 입은 것과 유사한 형태로 모델이 생성된 것을 관찰할 수 있었다. 앞선 과정들에서 생겼던 노이즈에 의해 완벽히 깨끗한 형태의 아바타는 아니었지만 만들어진 옷 모델과 인체 모델이 잘 합쳐진 것으로 보였다.



<Fig.7 옷과 인체가 합쳐져 완성된 아바타>

3.4 애니메이션 결과

세계적으로 인기를 끌었던 PSY 의 <강남 스타일>의 댄스를 애니매이션시켜 보았다. 동작은 원하는 대로 동작하였으나, 리듬감이나 결과 화질 및 노이즈 처리 등에서 한계점을 보였다.

https://youtu.be/GTrxM75C0hQ (잘 된 경우)

https://youtu.be/313U7uL5fls (잘 안된 경우)

<Fig. 8 강남스타일 춤 애니메이션>

4. 결론

키넥트 카메라를 통해 촬영한 이미지를 통해 자연스러운 아바타를 만들고 그것이 춤을 추는 형태로 움직일 수 있게 하는 기술을 연구하였다. 어느 정도 성공적인 아바타가 생성되고 애니메이션도 춤을 잘 따라 할 수 있게 만들어졌다. 그러나 랜더링 결과 면에서 많은 한계점을 보였다. 그 원인은 다음과 같다. 우선 본 연구에서 매칭되는 모델은 팔다리와 몸통을 모델링하지만 얼굴이나 손등은 제대로 모델링하지 못하는 문제점이 있다. 또한 이미지 분할 알고리즘이 정교도가 한두 픽셀 이상의 오차를 보이고 세그멘테이션에 사용하는 이미지의 해상도에 제약이 있어 텍스쳐의 정확도가 충분치 못하다. 특히 옷과 그림자의 색이 유사하다던가 옷과 배경의 색이 비슷한 경우 문제가 심한데 현재는 크로마키를 이용하면 개선이 가능할 수도 있다. 또한 3 차원 포인트 클라우드의 구조적인 노이즈 문제로 정합이나 의상 모델이 정교하지 못하다. 또한 촬영 중 사람의 신체는 물론이고 키넥트 카메라 또한 흔들림 없고 불편함 없이 촬영하는 촬영기법을 고안해내야 조금 더 깔끔한 결과를 기대해 볼 수 있을 것이다. 마지막으로는 실제 의상처럼 자연스러운 애니매이션이 되기 위해서는 물리엔진을 적용할 필요가 있다.

감사의 글

본 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1045391).

참고 문헌

1. 통계청 “스마트 기기 활용 시간_평일” (https://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1028451&conn_path=I2)

2. Chan, Caroline, "Everybody dance now." CVPR 2019.

3. C.Y. Weng, C. Brian, and I. Kemelmacher-Shlizerman. "Photo wake-up: 3d character animation from a single photo." CVPR 2019.

4. M. Loper, "SMPL: A skinned multi-person linear model." ACM transactions on graphics (TOG) 34.6 (2015): 1-16.

5. F. Bogo, "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," ECCV 2016.

6. R. A. Güler, N. Neverova, and K. Iasonas, "Densepose: Dense human pose estimation in the wild." CVPR. 2018.

7. K. Gong, X. Liang, Y. Li, Y. Chen, M. Yang and L. Lin, "Instance-level Human Parsing via Part Grouping Network", ECCV 2018.

8. H. Yang, J. Shi, & L. Carlone, “Teaser: Fast and certifiable point cloud registration,” IEEE Transactions on Robotics, 37(2), 314-333, 2020