

향상된 3D 모델링과 유사도 측정을 이용한 동작 학습 도움

Dance Helper with Improved 3D Modeling and Similarity

권동현, 임현호

(DongHyeon Kwon and HyunHo Lim)

요약 : 본 논문은 3D 모델링의 체형과 동작의 복원에서 성능을 향상시키고, 사용자가 학습하고자 하는 동작과 사용자가 직접 입력한 동작을 모두 3D 모델링하여 학습하고자 하는 자세와 본인의 자세를 자세히 관찰할 수 있도록 한다. 나아가 해당하는 동작에 대하여 유사도를 측정하여 얼마나 비슷한지 알려주고, 틀린 관절을 표시하고 올바른 동작이 되려면 어떻게 수정해야 하는지 직관적으로 표시하여 동작 학습에 도움을 주는 기법을 제안한다. 동영상에서 개별 프레임에 대해서 3D 모델링이 진행되기 때문에, 체형이 균일하지 않고 동작이 틀린 경우가 발생하게 된다. 제안하는 기법은 이러한 문제점을 보완하기 위해서 연속적인 프레임이라는 동영상의 특징을 이용한다. 전체 프레임의 3D 모델링 정보를 가지고 있다는 점을 이용하여, 체형을 균일하게 하고 틀린 동작을 올바르게 복원한다. 3D 모델링의 관절 별 동작 파라미터를 이용하여 모델과 사용자의 유사한 정도를 측정하고, 올바른 자세를 취할 수 있도록 한다. 실험 결과 체형이 균일하게 복원되었음을 확인할 수 있었고, 동작 복원을 통해 3D 모델링의 정확도의 향상이 이루어졌다. 제안하는 기법으로 향상된 3D 모델링과 관찰, 직관적 유사도 표현을 이용해 동작 학습에 도움을 준다.

키워드: 동영상, 3D 모델링, 유사도, 체형 복원, 포즈 복원, 관찰, 동작 학습 도움

Abstract : In this paper, we improve the performance in restoring body shape and pose of 3D modeling, and make it possible to observe the posture and the posture of the person to be modeled by 3D modeling of both the model and the user. We propose a technique to help learn motion by intuitively indicating how similar it is to measure, how to show the wrong joints, and how to move to get the correct posture. Since the 3D modeling is performed for individual frames in the video, the body shape is not uniform and the pose may be wrong. In order to compensate for this problem, the proposed method uses 3D modeling information of the entire frame because it is a video, so that the body shape is uniform and the wrong pose is correctly restored. Using each joint pose parameter of 3D modeling, measure the similarity between the model and the user and make it possible to take the correct pose. As a result of experiment, it was confirmed that body shape was restored uniformly, and 3D modeling accuracy was improved through pose restoration. The proposed method helps to learn motion by using improved 3D modeling and observation and intuitive similarity expression.

Keywords : video, 3D modeling, similarity, shape, pose, observation, motion learning help

1. 서론

최근 K-pop이 큰 인기를 끌며 안무 학습에 대한 수요가 꾸준히 증가하고 있다. 이런 안무와 같은 동작을 학습할 때에는 자세를 바르게 습득하는 것이 가장 중요하다. 본 논문에서는 이러한 동작을 영상으로부터 3D 모델링을 진행하여, 학습하고자 하는 동작을 자세히 학습할 수 있게 도와주는 기법을 소개한다. 이때까지 수많은 3D 모델링 연구에서 체형이 불균일하거나 동작이 잘못 복원될 수가 있었다. 또한, 두 3D 모델 포즈 간의 유사도를 비교 분석하는 연구는 거의 진행되지 않아 3D 모델링이 비교 대상과 사용자 모두에게 적용되는 경우는 없었다.

본 논문에서는 이러한 기존의 문제점과 없었던 점을 해결하기 위해 여러 기법을 사용하였다. 우선 인체의 3D 모델은 SMPL(Skinned multi-person

linear model)^[1] 모델을 사용하였다. 기존 2D 좌표 취득을 위해 많이 사용되었던 DeepCut^[2]보다 성능이 좋은 OpenPose^[3]를 SMPLify^[4](Keep it SMPL)와 결합하여 성능을 더욱 높이고, 체형 근사 알고리즘과 전후 프레임을 이용한 동작 복원 알고리즘을 통해 향상된 3D 모델링을 제공한다. 전체 프레임의 체형 파라미터의 분포를 분석하고, 가장 분포가 많은 구간에 해당하는 체형 파라미터들의 평균을 구하여 적용하는 방식으로 outlier를 제거하여 체형을 균일하게 하였다. 동작이 잘못 복원된 경우는 OpenPose의 confidence 값을 통해 추정하고, 틀린 부분은 전후 프레임을 이용하여 다시 복원하였다. 모델링은 OpenDR^[5]을 이용하였고, 사용자가 원하는 프레임을 자유롭게 관찰할 수 있도록 한다. 이와 함께 동작의 유사도와, 지시선을 제공하여 사용자들이 더 자세히 동작을 학습할 수 있을 것이다.

II. 기존 연구

기존의 3차원 인체 복원 연구에는 크게 2단계로 이루어져 있다. 2D 영상으로부터 인체의 2D 관절 좌표를 획득하고, 해당 관절 정보를 이용하여 3차원 모델로 복원하는 방식이다. 인체 정보는 자세, 체형 정보로 나뉘게 되며 이와 같은 정보는 많은 연구자가 공유한 데이터 셋을 이용하게 된다. 딥러닝의 발전과 컴퓨터 비전 분야에 관한 관심의 증가로 인하여 해당 분야의 연구가 활발히 이루어지고 있다.

2D 영상으로부터 인체의 2D 관절 정보를 얻는 연구는 굉장히 다양하게 존재한다. 이 전부터 다양한 기법들이 소개되었는데, 최근에는 딥러닝의 발전으로 CNN 기반의 기법이 소개되었다. 그리고 그중에서도 OpenPose라는 기법은 각 관절에 대해서 heatmap을 배치하여 검출하고 이를 연결하여 skeleton을 구성하게 된다. 이를 통해 한 사람에 대해서 관절 좌표를 알 수 있게 된다. 그리고 SMPLify에서는 위에서 획득한 2D 인체 관절 정보를 이용하여 3D 모델을 생성한다. 다양한 인체 데이터 셋을 이용하여 미리 학습한 정보를 토대로, 가장 적합한 3D 모델을 찾아내는 역할을 한다. OpenDR에서는 해당하는 3D 좌표를 이용하여 렌더링을 할 수 있게 도와준다.

하지만 2D 영상에서 3D 인체를 복원한다는 것에는 약간의 한계점이 존재하게 된다. 인체 일부본이 가려져 있거나, 영상에 노이즈로 제대로 된 정보를 얻지 못하는 경우에 발생한다. 대부분의 3차원 인체 복원 연구는 단일 영상에 대해 이루어지는 경우가 많아 이러한 문제점을 해결하기 힘든 어려움이 존재하였다. 그리고 사용하는 데이터 셋이 대부분 서양인이기 때문에 동양인을 모델링을 하면 체형 정보가 맞지 않는 경우 또한 발생하였다.

더불어 기존의 연구는 3D 모델링을 중점적으로 하였고, 사용자를 함께 3D 모델링하여 자세히 관찰하고, 유사도를 측정해 주는 연구가 전혀 없었다. 그리고 2D 관절 좌표를 비교하여 동작의 유사도를 측정하는 경우는 많이 있었지만, 동작의 학습을 도와주는 것에는 많은 연구가 이뤄지지 않았다. 그래서 사용자가 자신의 동작에서 어떤 부분을 얼마나 틀렸는지를 아는 것이 쉽지 않았고, 원하는 동작을 학습하는 것에서 어려움이 존재하였다.

III. 제안하는 기법 및 기술

본 논문에서는 동영상 3D 모델링은 한 사람에 대해서 진행한다고 가정한다. 본 논문에서 제안하는 기법은 동영상이 연속적인 프레임이라는 것을 이용하여, 단일 영상에서는 얻을 수 없는 정보들을 다른 프레임에 이용하여 얻어서 문제를 해결한다. 체형은 모두 같은 사람인 점을 중점적으로 활용하고, 동작은 동작의 연속성이 있음을 중점적으로 활용한다. 이를 통해 복원된 3D 모델링 결과를 이용해 모델과 사용자 간의 유사도를 측정한다. 마지막으로 위 결과들을 사용하기 편리한 GUI 프로그램으로 구현하여 사용자가 쉽게 사용하고 직관적으로 결과를 확인할 수 있도록 한다.

체형이 기존의 연구에서 균일하지 않았던 이유는 개별 프레임에 대해서 3D 모델링을 수행했기 때문이다. 이를 해결하기 위해 동영상 전체 프레임에 대해서 먼저 3D 모델링을 진행하여 전체 프레임의 체형 파라미터들을 획득한다. 그림 1의 좌측 그림은 전체 프레임의 체형 파라미터의 분포이다. 한 사람을 대상으로 3D 모델링을 하였음에도 값이 일정하지 않다. outlier를 제거하기 위해 -10.0부터 10.0까지 20개 구간을 1.0 간격으로 나누었다. 가장 많은 데이터를 가진 구간 내의 데이터들의 평균을 취하여 체형을 전체 프레임에 적용한다.

동작 또한 개별 프레임에 대해서 2D 관절 정보 획득과 3D 모델링을 수행했기 때문에 제대로 복원되지 않는 경우가 발생한다. 이러한 경우를 OpenPose의 confidence를 이용해 인지한다. 그 후에 동작의 연속성을 이용하여 전후 프레임의 동작 파라미터로 해당 프레임을 다시 복원한다. 복원 과정은 2D 관절 정보 획득 과정에서 한 번 수행하고, 3D 모델링 과정에서 다시 한 번 수행하여 정확도를 높일 수 있도록 한다. 그림 1의 우측 그림의 빨간 선은 3D 모델링에서의 우측 손목의 confidence의 분포, 파란 선은 우측 손목의 동작 파라미터의 분포이다. confidence의 분포를 보면 다른 값들에 비해 굉장히 낮은 값을 가지는 프레임이 존재함을 확인할 수 있다. confidence가 낮다는 것은 동작이 이상할 확률이 높음을 의미하므로, 해당 프레임의 전후 프레임의 confidence가 높은 경우 전후 프레임의 동작 파라미터의 평균치를 이상 프레임에 적용하여 동작을 복원하는 방식으로 진행한다.

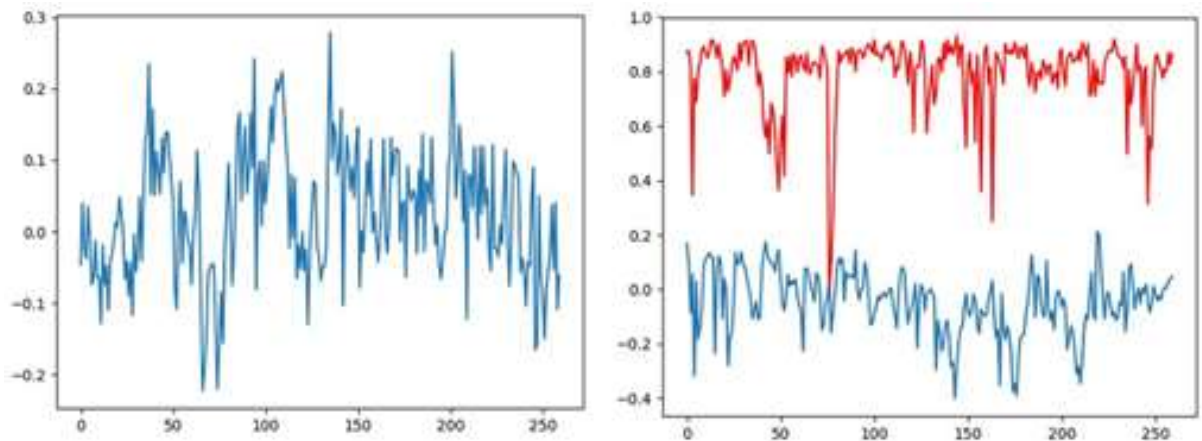


그림 1. 전체 프레임의 체형 파라미터 분포 (좌측), 전체 프레임의 우측 손목 confidence(빨간 선) 분포와 동작 파라미터(파란 선) 분포 (우측)

유사도 측정은 동작에 대한 유사도이므로 관절 별 동작 파라미터를 이용한다. 관절 별 유사도와 프레임 별 유사도를 각각 측정하고, 각 관절은 어깨-팔꿈치-손목이 팔을, 엉덩이-무릎-발목이 다리를 구성한다. 관절 별 유사도는 다음 수식과 같이 측정한다.

$$sim[i] = \frac{1}{(1 + \sum_{j=x}^z (m.pose[i][j] - m2.pose[i][j]))}$$

구한 관절 별 유사도를 통해 동작이 맞았는지 틀렸는지 아닌지를 판단한다. 비교적 움직임이 큰 팔은 0.8 미만으로 설정하고, 팔보다 움직임이 적은 다리는 0.6 미만으로 설정한다. 관절 별 유사도를 이용해 프레임 별 유사도를 측정하는 수식은 다음과 같다.

$$simF = \frac{\sum_{i=0}^{11} sim[i]}{12}$$

본 논문에서는 위와 같은 향상된 3D 모델링과 유사도 측정을 이용해 사용자에게 동작 학습을 돕는 것이 목표이다. 그러므로 최대한 사용자에게 친숙한 환경을 구축하고 직관적인 표현을 하기 위해 노력하였다. 사용자가 사용하기 편하도록 PyQT를 활용하여 GUI 환경을 구성하였다. 처음 사용하는 사용자라도 GUI 환경에서는 쉽게 사용할 수 있어 접근성이 좋기 때문이다. 또한, OpenDR을 이용하여 3D 렌더링을 진행하는데, 이때 파라미터를 조정하여 3D 모델을 다양한 시점에서 관찰할 수 있게 구현하였다. 이를 통해 2D 영상에서는 가려진 부분에 대한 동작도 관찰할 수 있다. 두 동작이 얼마나 일치하는지를 유사도로 나타낸다. 틀린 동작에 대해 원으로 표시하고, 올바른 동작을 위해서는 어느 방향으로 수정해야 하는 지를 화살표로 지시선을 표시하여 사용자가 직관적으로 알 수 있도록 한다.

IV. 실험 결과

본 논문의 실험은 Intel i5-9600k 4.6GHz CPU와 NVIDIA GeForce RTX 2060을 장착한 컴퓨터에서 수행하였다. 실험에 이용한 영상은 Youtube에서 취득하였고, 직접 촬영한 영상도 이용하였다. 동영상의 프레임 rate는 10fps으로 고정하여, 1초에 10개의 영상을 취득하여 사용했다. 한 프레임 당 약 2분의 수행 시간이 소요되었다. 멀티 프로세스를 이용하여 5장의 프레임을 동시에 수행하였고, 이를 통하여 한 프레임 당 약 25초의 수행 시간이 소요되었다.

그림 2는 동영상을 3D 모델링한 결과의 예시이다. 좌측 상단부터 우측 하단 방향으로 시간의 순서에 따라 영상이 진행된다. 실제 인체의 체형 및 동작과 3D 모델링의 결과가 거의 같고, 부드럽게 이어지는 것을 확인할 수 있다. 하지만 옆모습과 같이 인체의 일부분이 가려진 경우 정상적인 결과를 얻지 못하는 경우가 있는데, 이는 그림 3의 좌측에서 확인할 수 있다. 첫 번째 줄이 입력 영상이고, 두 번째 줄이 기존 기법에 따라 3D 모델링된 결과이다. 연속적으로 거의 동일한 자세를 취하고 있지만, 2, 4 번째 결과는 이상한 결과를 보이는 것을 알 수 있다. 그리고 세 번째 줄인 본 논문에서 제안한 기법을 이용한 모델링에서 정상적으로 복원한 것을 확인할 수 있다. 이것은 2, 4 번째에서 이상을 감지하고, 각각의 앞뒤 프레임에서 예측한 자세를 이용하여 복원해준 결과이다. 이는 실제 인체의 동작과 거의 유사한 것을 확인할 수 있다. 이를 통하여 연속적인 프레임에서 부드럽게 동작이 이어지는 결과를 얻을 수 있었다.



그림 2. 연속적인 프레임 영상을 3D 모델링한 결과

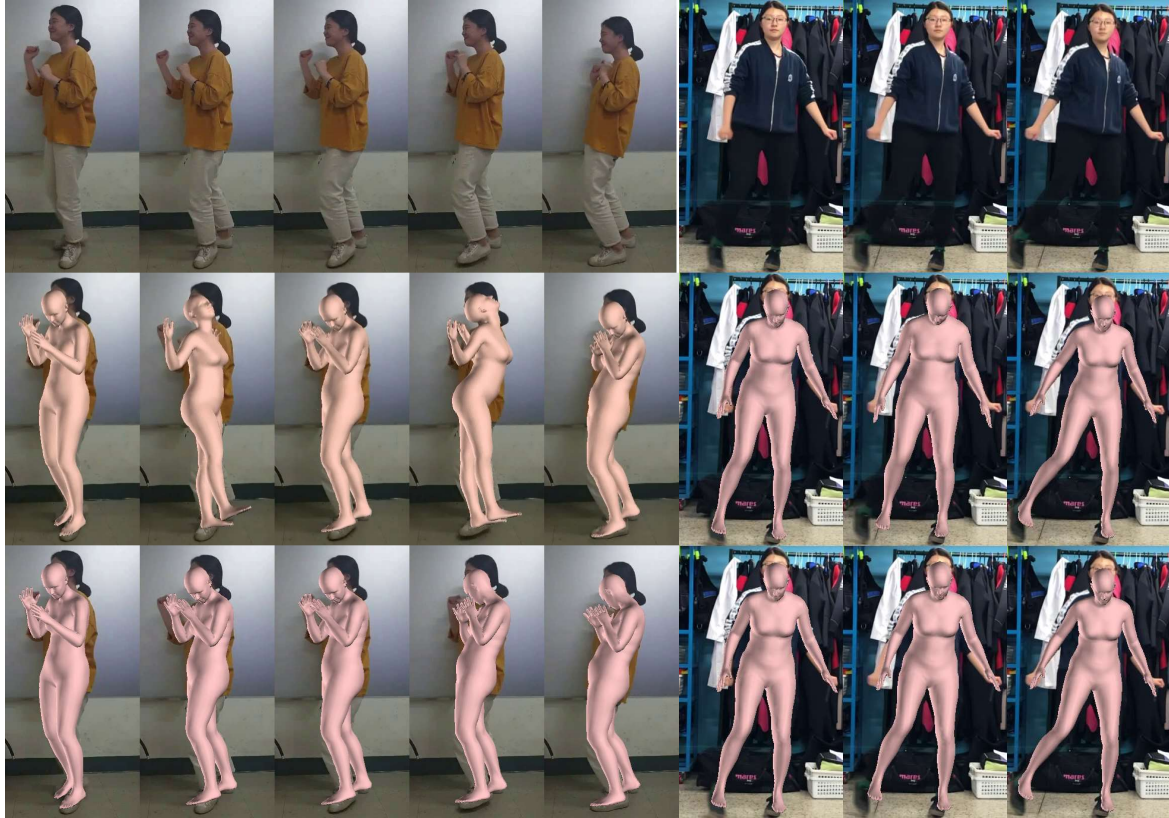


그림 3. 오류가 있는 프레임의 복원 결과. 좌측은 동작 오류가 있는 프레임의 보정 결과, 우측은 체형 오류가 있는 프레임의 보정 결과. 첫 번째 줄이 원본 영상, 두 번째 줄이 오류가 발생한 영상, 세 번째 줄이 오류를 보정한 영상

표 1. 전체 실험 영상에 대한 평균 에러 프레임의 수.

	Percentage of Error Frames
Existing Technique	14.69 %
Ours	13.91 %

체형이 불규칙하게 모델링 되는 결과는 그림 3의 우측에서 확인할 수 있다. 첫 번째 줄은 입력 영상이고, 두 번째 줄이 기존 기법에 따라 3D 모델링된 결과이다. 두 번째 프레임을 자세히 보면 양측에 있는 모델보다 체형이 더 크게 모델링된 것을 확인할 수 있다. 이것은 세 번째 줄인 본 논문에서 제안한 기법을 이용하여 균일한 체형으로 복원하였다. 이를 통하여 동작뿐만 아니라 체형까지 부드럽게 모델링을 진행할 수 있는 것을 확인할 수 있었다.

정량적인 평가를 위하여 Youtube에서 취득한 1인 안무 동영상에 이용하였다. 약 20개의 동영상을 이용하여 평가하였고, 평가 방식은 직접 각 프레임을 보고, 오류가 있는지 없는지를 판별하였다. 그리고 오류가 있는 프레임의 백분율을 이용하여 결과를 평가하였다. 기존의 기법은 약 14.7%의 오류율을 보였고, 본 논문에서 제안한 기법은 약 13.9%의 오류율로, 약 1%p의 성능 향상을 보여주고 있었다. 연속적으로 이상한 동작이나 인체 일부분이 가려진

경우가 입력되면, 동작을 복원하는 것이 어려워, 어느 정도의 한계는 분명히 존재하였다. 하지만 본 논문에서 제안한 기법이 연속적인 프레임의 동작을 보장하는 데에서 유의미한 효과가 있었고, 해당 방법을 더 발전시켜 더 좋은 결과를 얻어낼 수 있다는 것을 알 수 있었다.

본 논문의 궁극적인 목표는 사용자가 직접 프로그램을 이용하여 동작 학습을 할 수 있게 하는 것이었다. 그림 4와 같이 직접 구현한 GUI 프로그램을 이용하여, 사용자는 원하는 프레임으로 이동하여 자유롭게 동작을 관찰할 수 있다. 그리고 해당하는 프레임에서 자유롭게 시점을 변화하여 동작을 분석할 수 있게 한 모습을 확인할 수 있다. 더 나아가서 그림 5와 같이, 사용자가 따라 하고자 하는 영상과 사용자가 직접 입력한 영상을 비교하여 얼마나 비슷하게 따라 한 지 유사도를 분석해준다. 그리고 어떻게 동작을 수정해야 하는지 화살표로 지시선을 나타내어 올바른 동작을 취하기 위해 어떻게 해야 하는 지 확인할 수 있다.

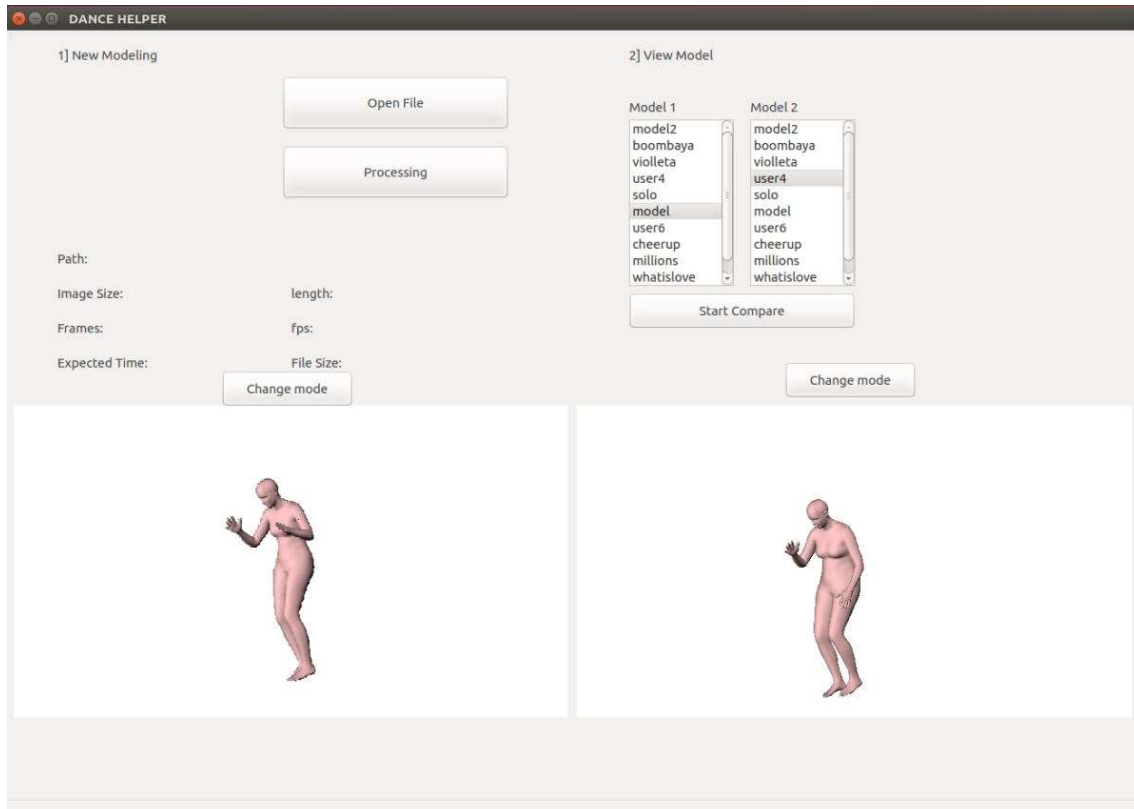


그림 4. 실제 사용자 GUI 프로그램을 사용한 결과.

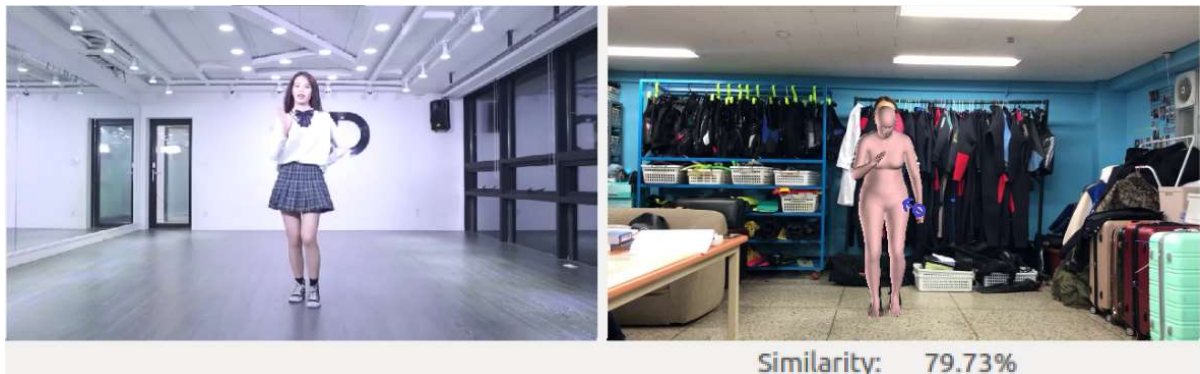


그림 5. 사용자가 따라 하고자 하는 영상 (좌측), 사용자가 직접 입력한 영상 및 유사도 측정 결과 (우측)

V. 결론

본 논문에서는 동영상에서의 3D 인체 모델링 복원의 보정 기법을 제시하였다. 한 동영상은 한 명에 대해 모델링한다는 가정에 따라, 여러 데이터를 이용하여 이상점을 제거하고, 가장 이상적인 평균값을 이용하여 인체 정보를 복원한다. 인체 일부가 가려졌을 때 모델링을 하는 데 있어서 정상적으로 수행되지 않는 경우가 발생한다. 이것을 연속적인 프레임이라는 동영상의 특징을 이용하여, 이상이 검출된 영상의 전후 정상적인 프레임에서 동작을 복원한다. 이러한 방식을 통해, 체형 및 동작의 오류가 있는 프레임에 대해서 정상적으로 복원한 결과를 얻을 수 있었다. 하지만 연속적으로 이상한 프레임이 검출되면, 정상적인 프레임에서 동작을 복원하는 방법이 잘 적용되지 않는다는 문제점이 발생했다. 하지만 본 논문에서 제안한 기법을 이용

하여 유의미한 발전이 있다는 것을 확인할 수 있었고, 이 방법을 발전시켜 더 좋은 결과를 얻을 수 있을 것이다.

사용하기 쉬운 GUI 프로그램을 구현하여, 사용자가 원하는 프레임을 선택하여 관찰할 수 있게 하였다. 원하는 프레임에서 시점을 자유롭게 변경하여 동작을 세밀하게 관찰하고, 쉽게 학습할 수 있게 하였다. 그리고 사용자가 따라 하고자 하는 영상과 사용자가 자신의 동작을 직접 입력하여 동작의 유사도를 분석해주었다. 그리고 화살표로 지시선을 나타내주어 어떻게 동작을 수정하는지 알려주는 방식을 통해, 동작을 학습하는 데에서 큰 도움을 줄 수 있게 되었다. 앞으로 수행 시간을 단축시켜 실시간으로 처리할 수 있게 된다면, 더 효율적으로 동작을 학습할 수 있을 것으로 기대할 수 있다.

참고문헌

- [1] M. Loper et al., “SMPL: A skinned multi-person linear model“, ACM Trans. on Graphics, vol. 34, no.6, Article No. 248, November 2015.
- [2] L. Pishchulin et al. “DeepCut:Joint subset partition and labeling for multi person pose estimation“, June 2016
- [3] Zhe Cao et al. “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields“, 14 April 2017.
- [4] F. Bogo et al. “Keep it SMPL: Automatic estimation of 3D human pose and shape from a single im- age“, October 2016.
- [5] M. Loper et al., “OpenDR: An approximate differentiable renderer,” September 2014
- [6] P. J. Rousseeuw et al., “Robust regression and outlier detection,” Vol. 1. New York: Wiley, 1987.
- [7] 한지수, 박인규, “분위 회귀 분석을 이용한 비디오로부터의 3차원 인체 복원,” 방송공학회 논문지 제24권 제2호, 2019. 3



권 동 현 (인하대 정보통신4)

2014년~현재 인하대학교 정보통신공공학과 학사과정 재학 중.

2020년 인하대학교 정보통신공학과 졸업 예정.

관심분야는 컴퓨터 비전, 인공지능, 웹 프로그래밍



임 현 호 (인하대 정보통신4)

2014년~현재 인하대학교 정보통신공공학과 학사과정 재학 중.

2020년 인하대학교 정보통신공학과 졸업 예정.

관심분야는 컴퓨터 비전, 인공지능