

3 차원 볼류메트릭 데이터 변형을 위한 선택적 시점에서의 3 차원 포즈 추정

이솔, 김지현, 박정탁, 박병서, *서영호

광운대학교

solee@kw.ac.kr, never4250@kw.ac.kr, jtpark@kw.ac.kr, bspark@kw.ac.kr,

*yhseo@kw.ac.kr

3D Pose Estimation from Selective View for 3D Volumetric Data Deformation

Sol Lee Ji-Hyun Kim Jung-Tak Park Byung-Seo Park *Young-Ho Seo

Kwangwoon University

요 약

본 논문에서는 선택적 시점에서의 2D 포즈 추정(pose estimation) 결과를 정합 하여 정확도 높은 3D 스켈레톤(skeleton)을 만들어 낸다. 여러 프레임의 3D 데이터를 10도 간격으로 36방향에서 투영한 뒤, 2D 포즈 추정 결과 신뢰도가 높은 시점에서의 결과만을 선별하여 3차원으로 정합 한다. 이때 사용하는 시점의 개수를 달리하며 정확도에 미치는 영향을 분석하여 실험적으로 정확도가 높은 최소의 시점 개수를 정하였다. 또한, 정합 한 3D 뼈대를 모션 캡처(motion capture) 센서와 비교하여 제안하는 알고리즘에 의해 3D 포즈 추정의 정확도가 향상되는 것을 확인했다.

1. 서론

최근 3 차원 영상 처리 기술의 연구가 활발해짐에 따라 3D 데이터의 수요가 증가하고 있지만, 그 용량이 너무 크다는 단점이 있다. 따라서 3D 데이터의 압축이 필수적이며, 본 논문에서는 그림 1 과 같이 비강체(non-rigid) 변환을 통해 다음 프레임의 3D 데이터를 생성하는 압축 방식에 필요한 고정밀 3D 포즈 추정 알고리즘을 제안한다. 강체(rigid) 변환이란 점들 사이의 간격이 변하지 않는 변환으로 평행이동과 회전 변환을 포함한다. 비강체 변환은 점들 사이의 간격이 변하는 변환으로 하나의 물체를 둘 이상의 부분으로 나누고 서로 다른 변환을 적용함으로써 수행할 수 있다. 3D 휴먼 데이터의 부분 분류 및 비강체 변환[1]은 3D 스켈레톤을 기준으로 이루어질 수 있으며, 그 정확도를 위해 정밀한 3D 포즈 추정의 방식이 필요하다.

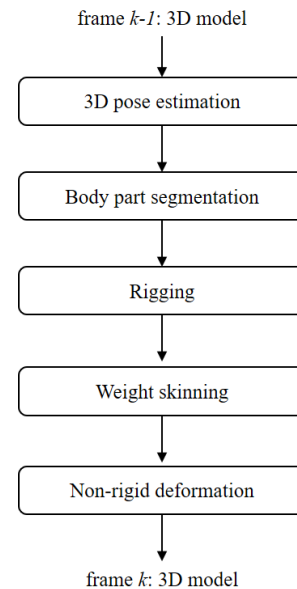


그림 1. 비강체 변환을 이용한 3D 데이터 압축 방식

2. 선택적 시점에서의 3 차원 포즈 추정

실험은 총 10 프레임의 볼류메트릭 촬영된 3D 데이터에 대해 다시점에서 투영(Projection) 이미지에서 2D 포즈를 추정한 결과를 기반으로 3D 스켈레톤을 생성된 데이터를 이용하였다. 이때, 각 관절마다 해당 관절이 잘 보이는 시점이 있고, 몸에 의해 가려지게 되는 시점이 있다. 그림 2.(a)는 모든 관절이 보이는 시점에서의 2D 관절 추정 결과이고, 그림 2.(b)는 오른쪽 팔다리가 몸통에 의해 가려져 오른 팔다리에 대한 2D 관절 추정이 불가능한 경우이다. OpenPose[2]는 가려지는 영역에 대해 추정하는 경우에 상대적으로 낮은 신뢰도를 제공한다.

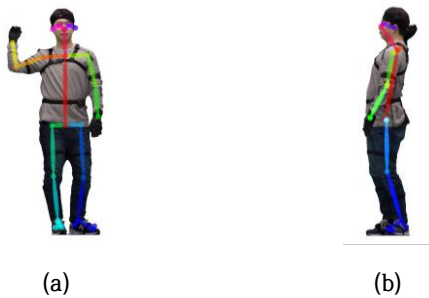


그림 2. (a) 모든 관절이 보이는 시점, (b) 일부 관절이 가려지는 시점에서의 2D 관절

OpenPose 가 제공하는 신뢰도를 내림차순으로 정렬하여 신뢰도 1 순위부터 8~32 순위까지의 결과를 이용하여 3D 관절을 생성한다. 이렇게 생성한 3D 관절을 모션 캡처 센서로부터 얻은 3D 관절과의 유클리드 거리차의 시간에 따른 표준편차를 구함으로써 정밀도를 측정한다.

3. 실험 결과

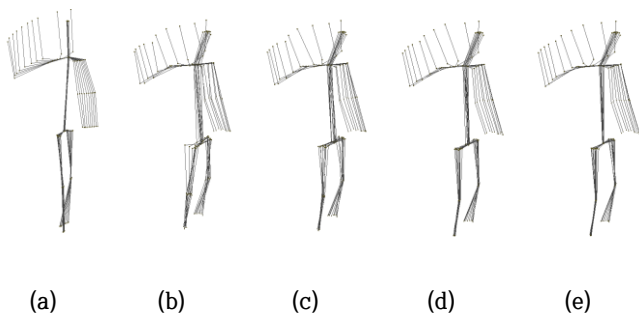


그림 3. (a) 모션 캡처 센서의 3D 관절, (b) 상위 8개, (c) 16개, (d) 24개, (e) 32개의 2D 관절로 생성한 3D 관절

그림 3 의 (a)는 모션 캡처 센서로부터의 3D 스켈레톤이고, (b)는 8, 16, 24, 32 순위까지의 2D 관절을 정합 하여 생성한 3D 스켈레톤의 결과이다. 모션 캡처 센서로부터 획득한 3D 스켈레톤을

기준으로 두고 비교해보면 상위 8 개의 시점을 이용해 정합 한 데이터는 골반과 무릎의 정밀도가 떨어진다. 그림 3.(d), (e)의 상위 24 개, 32 개의 시점을 정합 한 데이터는 둘의 차이를 육안으로 관찰하기 어려울 정도로 데이터가 유사한 것을 확인할 수 있다.

표 1.각 관절 별 정합 한 view 개수에 따른 TSD(Temporal Standard Deviation)와 관절 평균 TSD

	View Count / TSD			
	8	16	24	32
Before	16.164	12.378	10.714	9.834
After(Applying the Method)	12.046	10.451	10.353	9.721

표 1 은 알고리즘 적용 전과 후의 3D 포즈 추정에 사용된 시점의 개수 별 3D 스켈레톤과 모션 캡처 센서의 3D 스켈레톤을 비교한 오차를 표기하였다.

4. 결론

높은 신뢰도 값을 가지는 시점의 데이터를 이용해 3D 스켈레톤을 생성했을 때, 더 정밀한 3D 스켈레톤을 생성하는 경향을 보였고, 시점이 많아질수록 정밀도의 상승 폭이 줄어들며 수렴되는 결과를 보였다.

Acknowledgement

이 논문은 2021 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2018R1D1A1B07043220)

Reference

- [1] Morgenstern, W., Hilsmann, A., & Eisert, P., "Progressive non-rigid registration of temporal mesh sequences," *European Conference on Visual Media Production*, United Kingdom, London, pp. 1-10, 2019. doi: <https://doi.org/10.1145/3359998.3369411>
- [2] Z. Cao and G. Hidalgo and T. Simon and S. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 43, No.01, pp.172-186, Jan 2021. doi: <https://doi.org/10.1109/TPAMI.2019.2929257>