한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 29, No. 5, P. 11~20

트래커를 활용한 딥러닝 기반 실시간 전신 동작 복원

김현석[1] 강경원[2] 박강래[2] 권태수*

한양대학교 일반대학원 컴퓨터 소프트웨어학과

tomy8992@gamil.com, kang361973@gmail.com,gang31115@gmail.com, taesoobear@gmail.com

Deep Learning-Based Motion Reconstruction Using Tracker Sensors

Hyunseok Kim[1] Kyungwon Kang[2] Gangrae Park[2] Taesoo Kwon*

Dept. of Computer and Software, Hanyang University

요 약

본 논문에서는 손 동작을 포함한 전신 동작 생성이 가능하고 동작 생성 딜레이를 조절할 수 있는 새로운 딥러닝 기반 동작 복원 기술을 제안한다. 제안된 방법은 범용적으로 사용되는 센서인 바이브 트래커와 딥러닝 기술의 융합을 통해 더욱 정교한 동작 복원을 가능하게함과 동시에 IK 솔버(Inverse Kinematics solver)를 활용하여 발 미끄러짐 현상을 효과적으로 완화한다. 본 논문은 학습된 오토인코더(AutoEncoder)를 사용하여 트래커 데이터에 적절한 캐릭터 동작의 실시간 복원이 가능하고, 동작 복원 딜레이를 조절할 수 있는 방법을 제안한다. 복원된 전신 동작에 적합한 손 동작을 생성하기 위해 FCN(Fully Connected Network)을 사용하여 손 동작을 생성하고, 오토인코더에서 복원된 전신 동작과 FCN 에서 생성된 손 동작을 합쳐 손 동작이 포함된 캐릭터의 전신 동작을 생성할 수 있다. 앞서 딥러닝 기반의 방법으로 생성된 동작에서 발 미끄러짐 현상을 완화시키기 위해 본 논문에서는 IK 솔버 를 활용한다. 캐릭터의 발에 위치한 트래커를 IK 솔버의 엔드이펙터(end-effector)로 설정하여 캐릭터의 발 움직임을 정확하게 제어하고 보정하는 기술을 제안함으로써, 생성된 동작의 전반적인 정확성을 향상시켜 고품질의 동작을 생성한다. 실험을 통해, 본 논문에서 제안한 딥러닝 기반 동작 복원에서 정확한 동작 생성과 사용자 입력에 따라 프레임 딜레이 조정이 가능함을 검증하였고, 생성된 전신 동작의 발미끄러짐 현상에 대해 IK 솔버가 적용되기 이전 전신 동작과 비교하여 보정에 대한 성능을 확인하였다.

Abstract

In this paper, we propose a novel deep learning-based motion reconstruction approach that facilitates the generation of full-body motions, including finger motions, while also enabling the online adjustment of motion generation delays. The proposed method combines the Vive Tracker with a deep learning method to achieve more accurate motion reconstruction while effectively mitigating foot skating issues through the use of an Inverse Kinematics (IK) solver. The proposed method utilizes a trained AutoEncoder to reconstruct character body motions using tracker data in real-time while offering the flexibility to adjust motion generation delays as needed. To generate hand motions suitable for the reconstructed body motion, we employ a Fully Connected Network (FCN). By combining the reconstructed body motion from the AutoEncoder with the hand motions generated by the FCN, we can generate full-body motions of characters that include hand movements. In order to alleviate foot skating issues in motions generated by deep learning-based methods, we use an IK solver. By setting the trackers located near the character's feet as end-effectors for the IK solver, our method precisely controls and corrects the character's foot movements, thereby enhancing the overall accuracy of the generated motions. Through experiments, we validate the accuracy of motion generation in the proposed deep learning-based motion reconstruction scheme, as well as the ability to adjust latency based on user input. Additionally, we assess the correction performance by comparing motions with the IK solver applied to those without it, focusing particularly on how it addresses the foot skating issue in the generated full-body motions.

키워드: 모션 캡처, 딥러닝, 바이브 트래커, 실시간 모션 복원, 발 미끄러짐 현상
Keywords: Motion Capture, Deep Learning, Vive Tracker, Real-time Motion Reconstruction, foot skating

*corresponding author: Taesoo Kwon / Department of Computer Science, Hanyang University Graduate School (taesoobear@gmail.com)

Received : 2023.10.04./ Review completed : 1st 2023.11.03. / Accepted : 2023.11.10.
DOI : 10.15701/kcgs.2023.29.5.11
ISSN : 1975-7883(Print)/2383-529X(Online)

1 서론

현대 디지털 엔터테인먼트와 가상현실(VR) 분야의 급속한 진화는 현실감 있는 모션 캡쳐 시스템에 대한 수요를 증가시켰다[1]. 모션 캡쳐 시스템은 영화, 비디오 게임, 가상현실 체험 등에서 현실적이고 자연스러운 캐릭터 동작을 생성하는 데 중요한 역할을 한다. 고품질의 동작을 생성하기 위해 기존 방법들에서는 다수의 트래커와 카메라 센서들을 활용하여 정확하고 자연스러운 동작을 복원하였다. 하지만 모션 캡쳐 장비들은 대부분 시장에서 높은 가격대를 형성하고 있어 범용적이지 않다는 한계가 있다. 최근 온라인 가상 비디오 컨텐츠가 범용화 되면서 소수의 저비용 센서를 활용하여 동작을 복원하는 연구들이 활발히 연구되고 있다[2][3]. 저비용 모션 캡쳐 장비는 정확성이 부족하고 노이즈가 있어 생성된 동작이 부자연스러울 수 있다[4].

최근에는 이러한 문제를 해결하기 위해 딥 러닝 기술을 접목하여 자연스러운 동작 복원이 가능한 방법들이 제시되고 있다. 그러나 소수의 센서를 활용하여 동작 복원을 수행하는 딥러닝 기반 방법들은 여전히 실제 움직임과 다르게 발이 지면에서 미끌어지는 발 미끄러짐 현상이 발생할 수 있다. 또한 대부분의 딥러닝 기반의 방법들은 손 동작이 제외된 전신 동작을 복원하는 데 중점을 두어 캐릭터의 몸의 동작만을 반영하고 손 동작 표현이 제한된다. 손 동작의 부재는 전체적인 동작의 풍부성과 표현력을 제한할 수 있다. 또한 기존의 딥러닝 기반의 방법들은 다수의 프레임에서 얻은 트래커 데이터를 활용하기 때문에 고정된 딜레이를 가지는 동작을 복원하여 실시간 동작 복원에 어려움이 있을 수 있다. 이러한 문제를 완화하기 위해 소수의 센서를 활용하는 동작 복원에서 정확한 전신 동작과 자연스러운 손 동작을 생성하며, 사용자의 의도에 맞게 타이밍을 조절하고 더 나아가 물리적으로 안정된 동작을 보장하는 것은 중요한 과제로 인식되어지고 있다.

본 논문에서는 앞서 소개한 한계점들을 보안하는 동작 복원 방법을 제안한다. 제안된 방법은 손 동작을 포함한 전신 동작을 생성하며 동작 생성 딜레이를 조절할 수 있는 딥러닝 기반 동작 복원 시스템 개발을 목표로 한다. 먼저 6개의 바이브 트래커에 대한 정확한 위치,회전 값을 촬영하고, 촬영된 트래커 데이터를 학습된 오토인코더의 입력으로 사용하여 저차원 레이턴트(latent) 벡터로 인코딩한다. 그리고 앞서 생성된 레이턴트 벡터를 디코딩하여 전신 동작을 복원한다. 오토인코더에서 복원된 동작은 사용자의 입력에 따라 최소 1 에서 10 프레임까지 딜레이 조절가능하다. 이를 통해 사용자는 빠른 반응 속도와 동작 품질 사이의 트레이드 오프를 적절하게 선택할 수 있다. 그 후, 복원된 전신 동작을 입력으로 사용하여 학습된 FCN를 통해 적절한 손 동작을 생성한다. 앞서 생성된 전신 동작과 손 동작을 결합하여 캐릭터의 최종 전신 동작을 생성한다. 마지막으로, 기구학적 동작 생성 방법에서 발생할 수 있는 발 미끄러짐 현상을 완화하기 위해 IK 솔버를 통해 캐릭터의 발 움직임을 더욱 정교하게 제어하여 생성된 동작을 보정한다. 그 결과 실시간 동작 복원이 가능하며 더욱 정교하고 자연스러운 캐릭터 전신 동작을 생성할 수 있다. 본 논문에서 제안된 방법을 요약한 내용은 다음과 같다.

- 딥러닝을 활용하여 정확한 전신 동작과 자연스러운 손 동작 생성 가능
- 딜레이 조절이 가능하여 사용자의 의도에 맞는 타이밍에 동작 생성 가능
- IK 솔버를 사용한 보정을 통해 생성된 동작에서 발 미끄러짐 현상 감소

본 논문의 구성은 다음과 같다. 2장에서는 관련된 연구들에 대해 소개하고 3장에서는 본 논문에서 제안하는 방법에 대해 상세히 설명한다. 4장에서는 제안한 방법에 대한 실험 및 평가를 수행하고 5장에서 결론을 맺는다.

2 관련 연구

2.1 비전 기반의 동작 복원

다양한 센서를 이용한 동작 복원 기술은 컴퓨터 그래픽스와 컴퓨터 비전 분야에서 활발히 연구되고 있다. 본 장에서는 다양한 분야에서 활용 가능한 실용적인 기술들을 개발하고 있는 비전 기반의 동작 복원 연구들을 소개한다. 다중 카메라 시스템, 딥러닝, 실시간 처리 등의 발전을 통해 정확하고 자연스러운 동작 복원이 가능해지며, 이는 더욱 현실적이고 효과적인 가상 및 증강현실을 구현하는데 기여하고 있다[5]. Shafaei et al. 은 마커나 센서 없이 다중 depth 카메라만을 사용하는 딥러닝 기반 동작 복원 방법을 제안하였다[6]. 제안된 방법은 이미지 분할(segmentation) 기술을 활용하여 각 신체의 부위를 구분하고 커리큘럼 학습(curriculum learning)을 사용하여 모션 트래킹하는 모델을 만들었다. Mehta et al. 은 단일 RGB 이미지를 이용한 동작 복원 방법을 제안하였다[7]. 제안된 방법은 CNN 기반의 방법으로 향상된 CNN supervision 기술을 사용하여 기존 연구들보다 더 좋은 성능을 보였다. 하지만 제안된 방법을 통해 생성된 동작은 물리적 현실성을 보장하지 않는다. Zou et al. 은 단일 RGB 이미지를 활용한 동작 복원 시스템을 제안하였다[8]. 제안된 방법은 CNN 모델에 발 접촉에 관련된 부분이 추가된 모델을 형성하여 발 미끄러짐 현상이 완화된 동작 복원 시스템을 제안하였다. Lu et al. 은 depth 카메라를 이용해서 일반적인 신체 모션 트래킹뿐만 아니라, 얼굴 표정과 손의 자세까지 복원하는 방법을 제안하였다[9]. 하지만 앞서 소개한 비전 기반의 방법들의 경우 카메라 센서를 사용하기 때문에 조명, 배경, 가려짐 등의 요소로 인해 캡쳐의 정확도가 민감하고, 다중 카메라 시스템을 활용하는 방법의 경우 많은 양의 데이터

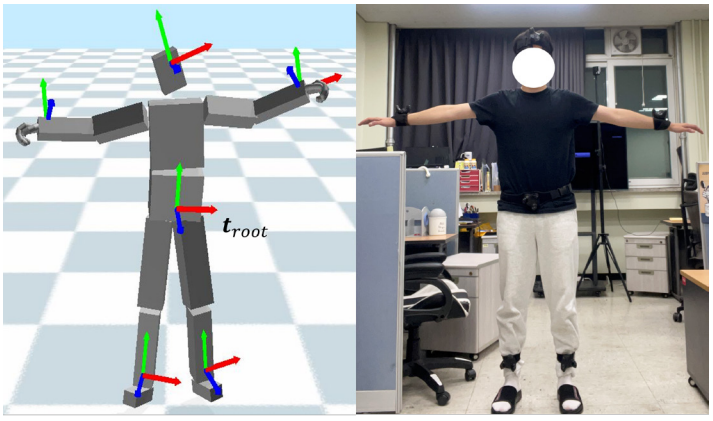
를 처리해야함으로 계산 복잡성을 증가시켜 동작 복원시 지연 문제를 야기시킬 수 있다.

2.2 마커 센서 기반의 동작 복원

지난 수십 년 동안 마커 센서를 활용하여 인체의 움직임을 감지하고 분석하는 기술이 활발히 연구되었다. 인체의 동작을 복원하기 위해 광학 트래커, 관성 측정 장치, 전자 자기장 센서 등 다양한 마커 센서들을 사용하여 사람의 동작을 측정하였다. 최근에는 이러한 다양한 마커 센서들과 딥 러닝 기술을 결합한 동작 복원 방법들이 활발히 연구되어지고 있는데, 이러한 방법들은 기존 방법들에 비해 센서 노이즈에 강건하고 자연스러운 움직임을 재현하거나 분석할 수 있다는 장점이 있다[10]. 마커 센서를 활용하는 대부분의 연구들은 6개 이하의 마커 센서를 활용하여 동작 복원 시스템을 구성하고 있다. Huang et al. 의 연구에서는 6개의 IMU 센서를 사용하고, RNN 기반의 신경망을 구성하여 새로운 구조의 딥 러닝 기반의 동작 복원 시스템을 제안하였고, 이를 통해 정확한 실시간 동작 복원이 가능한 시스템을 구현하였다[11]. YI et al. 은 6개의 IMU 센서를 사용한 딥 러닝 기반의 동작 복원 방법을 제안하였는데, 제안된 방법은 새로운 구조의 bi-RNN 기반 신경망 모델을 사용하여 정확하고 동작 복원에서 중요하게 여기는 캐릭터의 골반 조인트의 위치와 방향까지 예측이 가능한 방법을 제안하였다[12]. Kim et al. 은 5개의 IMU 센서와 HMD(Head Mounted Display)를 함께 사용한 방법을 제안하였다[13]. 이 방법은 기존 연구에서 사용된 bi-RNN 기반 신경망 모델과 새로운 convLSTM 신경망을 사용하여 빠르고 정확한 동작이 복원되는 방법을 제안하여 실용성을 평가하였다. 이와 같이 6개의 마커 센서를 사용하여 동작 복원 시스템을 구성한 연구도 있는 반면 실용성을 높이기 위해 센서의 개수를 줄여 동작 복원하는 방법에 대한 연구도 활발히 진행되고 있다. Yang et al.의 연구에서는 4개의 상체 관절에 트래커 신호만을 사용하여 하체의 자세를 실시간으로 예측하는 딥러닝 기반의 동작 복원 방법을 제안하였다[14]. 제안된 방법에서는 트래커의 속도와 하체 동작간의 상관관계를 모델링하여, 다양한 체형 및 비율에 강건한 동작 복원 방법을 제안하였다. 최근에는 이보다 더 적은 수의 마커 센서를 사용하는 동작 복원 방법이 개발되고 있는데, 그 중 AHUJA et al. 은 HMD 와 2개의 컨트롤러 만을 사용한 동작 복원 기술을 제안하였다[15]. 이 기술은 기존의 모션 캡쳐 데이터베이스를 모션 저장소로 사용하며, 데이터베이스 내에 유사한 시공간의 동작을 찾아 생성하여 더욱 몰입감있는 아바타의 동작을 생성할 수 있는 동작 복원 시스템을 제안하였다. 하지만 이전에 소개한 연구들은 물리법칙이 적용되지 않은 기구학적 동작 생성 방법을 사용하기 때문에 떨림이나 발 미끄러짐 현상이 발생할 수 있다. 이를 완화하기 위해 다른 연구들은 물리 시뮬레이터 상에서 토크 기반으로 동작을 생성하는 동작 복원 시스템을 제안하였다. Yi et al. 은 6개의 IMU 센서를 사용한 새로운 동작 복원 시스템을 제안하였다[16]. 제안된 방법은 기구학적 동작을 생성하는 모델과 각 관절의 회전 값과 위치에 대한 dual PD 제어기를 활용하여 물리 시뮬레이터 기반의 동작 복원 시스템을 제안하였다. Winkler et al. 은 1개의 HMD 와 2개의 컨트롤러를 활용한 물리 시뮬레이터 상의 동작 복원 시스템을 제안하였다[17]. 이 시스템은 강화학습과 PD 제어기를 활용하여 다른 물체와 상호작용하는 동작을 물리 시뮬레이션을 통해 복원한다. 이처럼 동작 복원에 관한 연구들이 다양하게 진행되어지고 있지만 대부분의 연구들의 경우 손 동작을 제외한 전신 동작 생성에만 중점을 두고있다. 또한 기존의 동작 복원 방법들은 소수의 트래커 센서를 사용하여 모션을 생성하기 때문에 여러 프레임의 트래커 데이터를 활용하여 고정된 딜레이가 생기게 된다. 본 논문에서는 대표적인 저가격 센서인 바이브 트래커와 딥러닝 기술을 결합하여 손 동작이 포함된 전신 동작을 생성하는 동작 복원 방법을 제안한다. 본 논문에서 제안되는 방법은 복원되는 동작의 딜레이 조절이 가능하고 기존의 방법들보다 비교적 적은 트래커 데이터 버퍼를 사용하여 빠르게 동작을 생성할 수 있다. 또한 IK 솔버를 생성된 동작에 적용하여 저비용 센서를 사용한 기구학적 동작 생성 방법에서 빈번히 발생하는 흔들림이나 발 미끄러짐 현상을 완화하는 동작 복원 시스템을 제안한다.

3 딥러닝 기반 동작 복원

3.1 캐릭터 모델



(a) (b)

Figure 1: (a) Character model used in this research, (b) The capture subject wearing 6 vive trackers.

Figure 1(b)와 같이 본 논문에서는 총 6개(손목, 발목, 골반, 머리)의 트래커 데이터를 사용한다. Figure 1(a) 는 실제 트래커 데이터를 시뮬레이션 환경으로 변환하여 나타낸 좌표계를 나타낸다. 하지만 각 관절의 좌표와 트래커의 실측 데이터는 약간의 오차가 존재한다. Figure 1(a)의 각 관절의 좌표는 실제 트래커 데이터의 차이만큼 오프셋을 적용하여 차이를 최소화

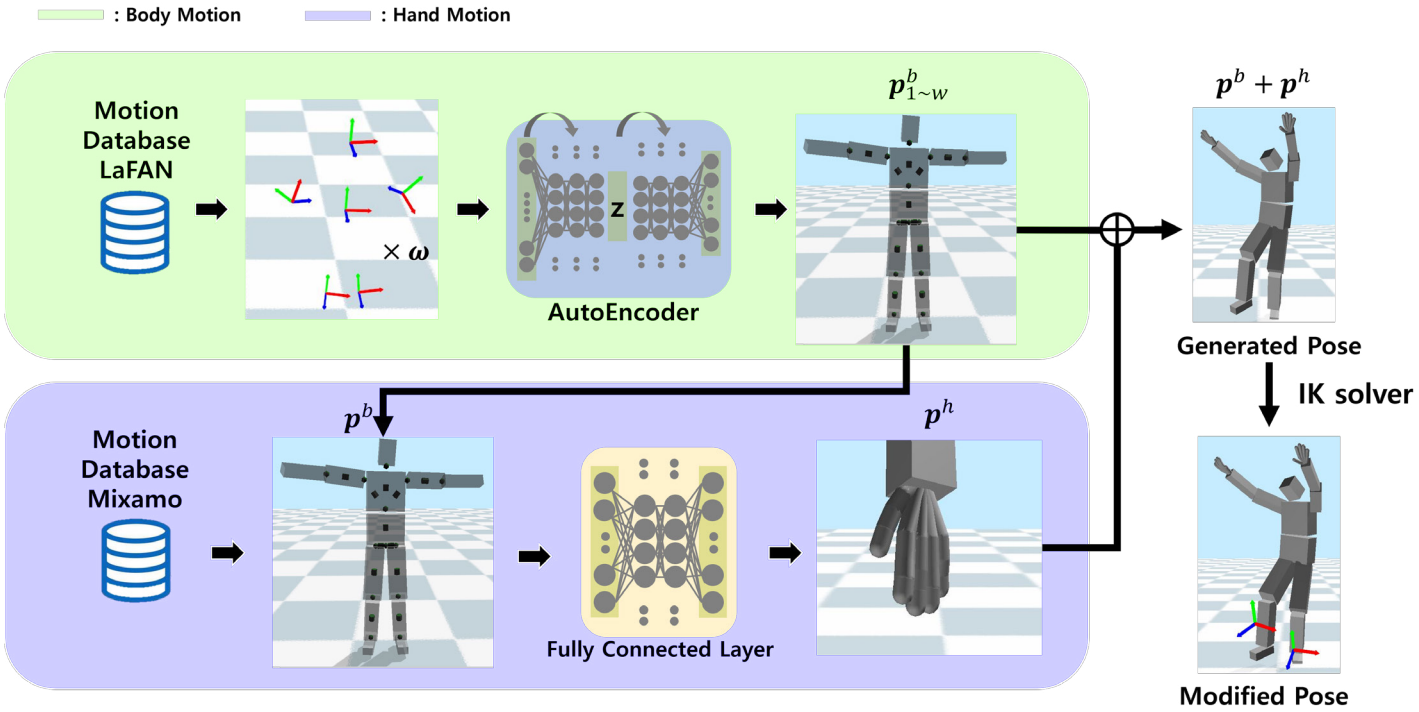


Figure 2: Overview of Motion Reconstruction System.

한 결과이다. 캐릭터는 손에 56 DOFs(Degree of Freedom)을 전신 동작에 26 DOF 를 가지고 있으며 root 관절인 $\boldsymbol{t}_{root}$ 에 7 DOF 합하여 총 89 DOF를 가지고 있다.

3.2 학습 데이터

전신 동작 복원 모델은 6개의 트래커 데이터를 입력으로 활용하여 동작을 복원한다. 전신 동작 모델 학습시 대용량 모션 데이터셋인 Ubisoft 사의 LaFAN 의 모션 데이터셋에서 트래커 데이터의 상응하는 좌표를 사용하여 학습을 진행한다[18]. 전신 동작 복원 모델을 학습하기 위해 사용되어지는 입력 데이터는 각 트래커의 위치 $\boldsymbol{t}^p \in \mathbb{R}^3$, 회전 값 $\boldsymbol{t}^r \in \mathbb{R}^6$, 선속도 $\dot{\boldsymbol{t}}^p \in \mathbb{R}^3$ 그리고 각속도 $\dot{\boldsymbol{t}}^r \in \mathbb{R}^3$ 이다. 위치와 회전 값은 Figure 1(a) 에서 $\boldsymbol{t}_{root}$ 좌표를 바닥으로 투영한 좌표계에 상대적으로 표현하여 학습에 사용된다. 회전 값의 경우 tangent normal 방법을 활용하여 6차원의 벡터로 표현하여 학습에 사용한다[19]. 캐릭터의 전신 동작 데이터 $\boldsymbol{p}^b \in \mathbb{R}^{26}$ 는 각 관절의 회전 값으로 표현한다. 최종적으로 본 논문에서 한 프레임에 해당하는 트래커 데이터 $\boldsymbol{t} \in \mathbb{R}^{90}$ 는 수식 (1) 과 같이 표현할 수 있다.

$$\boldsymbol{t} = \{\boldsymbol{t}^p_1, \boldsymbol{t}^r_1, \dot{\boldsymbol{t}}^p_1, \dot{\boldsymbol{t}}^r_1 ...., \dot{\boldsymbol{t}}^p_6, \dot{\boldsymbol{t}}^r_6\}. \tag{1}$$

손 동작을 생성하는 모델은 캐릭터의 전신 동작 데이터 $\boldsymbol{p}^b$ 를 네트워크의 입력으로 사용하여 손 동작 $\boldsymbol{p}^h \in \mathbb{R}^{56}$ 을 생성한다. 손 동작 모델 학습 시 손 동작까지 포함되어있는 MIXAMO 데이터를 사용하여 학습한다. 손 동작 데이터 $\boldsymbol{p}^h$ 의 경우 캐릭터의 전신 동작 데이터와 같이 손을 구성하는 각 관절의 회전 값으로 표현한다.

3.3 전신 동작 복원 모델

전신 동작 복원을 위해 본 논문에서는 오토인코더 네트워크를 활용한다[20]. 오토인코더는 입력 데이터를 저차원인 레이턴트 벡터(z)로 축소하여 표현하고 이를 다시 복원하여 입력 데이터와 같은 데이터를 생성하는 네트워크이다. 주로 데이터의 특징을 표현하고 차원축소를 위해 많이 활용되어진다. 본 논문에서는 오토인코더를 활용하여 윈도우 크기 w 개의 프레임에 해당하는 트래커 데이터 $\boldsymbol{t}_{i\sim i+w}$ 를 입력으로 사용하여 레이턴트 벡터(z) 차원에 인코딩하고 레이턴트 벡터(z)를 디코딩하여 전신 동작 $\hat{\boldsymbol{p}}^b_{i\sim i+w}$ 를 생성한다.

인코더 네트워크는 4개의 층을 가지고 있다. 은닉층은 1024개의 노드를 가지고 있고 입력 데이터가 잔차 연결 방식으로 각 은닉층에서 합쳐지는 구조이다. 인코더 네트워크의 입력으로는 w 개 프레임 트래커 데이터 $\boldsymbol{t}_{i\sim i+w} \in \mathbb{R}^{w\cdot 90}$ 가 사용되어 300 차원의 레이턴트 벡터(z)가 출력된다. 인코더 네트워크를 통해 트래커 데이터의 특징을 추출하고 차원 축소로 일반화하여 레이턴트 벡터(z)에 정의할 수 있다. 트래커 데이터의 특징이 정의된 레이턴트 벡터(z)를 활용하여 디코더에서 캐릭터의 자연스럽고 정확한 동작을 복원하도록 돕는다.

디코더 네트워크는 4개의 층을 가진 네트워크이고, 각 은닉층은 512 개의 노드를 가지고 있다. 디코더 네트워크의 입력은 앞서 인코더 네트워크에서 출력된 레이턴트 벡터(z)이고 출

력으로는 w 프레임의 전신 동작 $\hat{\boldsymbol{p}}^b_{i\sim i+w}$ 이 출력된다. 인코더 네트워크의 출력인 레이턴트 벡터(z)가 잔차 연결 방식으로 각 층에서 합쳐지는 구조로 이루어져있다. 여러 프레임의 전신 동작 복원을 통해 원하는 타이밍의 동작을 복원하고, 그에 따른 딜레이를 조절하여 최적의 결과를 얻을 수 있다. 본 논문에서는 윈도우 크기를 10으로 ($w = 10$) 설정한다. 인코더와 디코더 네트워크는 은닉층에서 ELU(Exponential Linear Unit) 활성화 함수를 사용한다. Figure 2에서 확인할 수 있듯이 실시간 동작 복원시 오토인코더에서 출력된 결과는 다시 손 동작 생성 모델의 입력으로 들어가는 방식으로 구성된다. 전신 동작을 생성하는 모델의 손실 함수 수식 (2)는 모션 데이터셋의 예제 동작 $\boldsymbol{p}^b_{i\sim i+w}$ 과의 MSE(Mean Squared Error)로 학습을 진행하였다:

$$loss_{fullbody} = \left\|\boldsymbol{p}^b_{i\sim i+w} - \hat{\boldsymbol{p}}^b_{i\sim i+w}\right\|_2^2. \tag{2}$$

이를 통해 트래커 데이터 $\boldsymbol{t}_{i\sim i+w}$ 에 따라 캐릭터의 적절한 전신 동작 $\hat{\boldsymbol{p}}^b_{i\sim i+w}$ 을 생성 할 수 있다.

3.4 손 동작 생성 모델

손 동작을 생성하는 모델의 경우, 4개의 층으로 구성되어있는 FCN 네트워크를 사용한다. 각 은닉층은 64개의 노드로 구성되어 있고, 입력으로는 캐릭터의 전신 동작 데이터를 사용하고 손 동작에 해당하는 데이터가 출력된다. 본 논문에서는 보다 정확한 손 동작을 생성하기 위해 3 프레임의 전신 동작 데이터를 입력 데이터로 사용한다. 예를 들어 학습 데이터 중 t 프레임의 손 동작을 생성하기 위해 MIXAMO 데이터셋에서 $t-1$ 에서 $t+1$ 시점의 전신 동작인 $\boldsymbol{p}^b_{t-1\sim t+1}$ 을 입력으로 사용하여 t 프레임의 손 동작 $\boldsymbol{p}^h_t$ 을 생성한다. 여러 프레임의 전신 동작을 입력으로 사용하여 전신 동작의 흐름과 손 동작간의 연관성을 파악할 수 있다. 손 동작 모델에서 은닉층의 활성화함수로 ELU 함수를 사용하였고, 손실함수로 모션 데이터셋의 손 동작 $\boldsymbol{p}^h$ 과 생성된 동작 $\hat{\boldsymbol{p}}^h$ 의 MSE로 수식 (3) 을 사용하였다:

$$loss_{hand} = \left\|\boldsymbol{p}^h - \hat{\boldsymbol{p}}^h\right\|_2^2. \tag{3}$$

이를 통해 3 프레임의 전신 동작의 적합한 손 동작을 생성할 수 있다.

3.5 IK 솔버

전신 동작을 복원하는 오토인코더와 손 동작을 생성하는 FCN 을 통해 트래커 데이터에 적절한 캐릭터 동작 생성이 가능해졌다. 하지만 기구학적 동작 생성 방법으로 생성된 동작은 발 미끄러짐 과 같은 현상이 발생할 수 있다. 이러한 경우 물리적으로 불안정한 동작을 생성할 수 있어 사용자의 몰입감을 감소시킬 수 있다. 이러한 문제를 해결하기 위해 본 논문에서는 발의 정확한 제어를 위해 발 관절의 트래커 데이터를 엔드이펙터로 사용하여 역기구학(IK)를 푸는 방법으로 캐릭터의 하반신 동작을 제어한다. 우리는 IK과정에서 네트워크가 생성한 자세가 많이 변형되는 것을 막기 위하여, 무릎 관절의 댐핑이 포함된 IK 솔버[21]를 활용하여 생성된 동작에 대해 보정을 수행한다. IK 솔버는 수식 (4) 을 이용하여 캐릭터 무릎 관절이 너무 빠르게 펴지는 것을 막는다:

$$\theta = \theta_0 + \int_{\theta_0}^{\theta+\Delta\theta} f(x)\,dx, \tag{4}$$

여기서 θ_0 은 현재 캐릭터의 무릎 관절의 회전 값을 나타내고, $\Delta\theta$ 는 현재 캐릭터의 관절의 회전 값과 트래커 데이터기반으로 계산되어진 관절의 회전 값의 차이를 나타낸다. $f(x)$ 는 무릎이 완전히 펴질 수록 가중치가 낮아지도록 하는 함수로 아래와 같이 정의된다:

$$f(x) = \begin{cases} 1, & x < \rho \\ \alpha(\frac{x-\rho}{\pi-\rho}), & \text{otherwise} \end{cases} \tag{5}$$

여기서 ρ는 사용자가 정하는 회전 값 임계치를 나타내는데 본 논문에서는 130도로($\rho = 130\,^\circ$) 설정하여 사용한다. 수식 (5)에서 사용되는 $\alpha(t)$ 는 아래와 같다:

$$\alpha(t) = t^3(0.5t - 1) + t. \tag{6}$$

이러한 수식을 통해 트래커 데이터 기반으로 캐릭터의 발 움직임을 더욱 정확히 제어할 수 있고, 발 미끄러짐 현상을 완화할 수 있다.

4 실험 및 평가

본 장에서는 논문에서 제안한 동작 복원에 대한 유효성과 성능을 입증하기 위해 다양한 실험을 수행했다. 동작 복원 방법에 대한 개발 및 실험 수행을 위해 AMD Ryzen 5 5600X 6-Core Processor 3.7GHz, 32GB RAM, NVIDIA GeForce RTX 3070 Ti 로 구성된 PC를 사용하였다. 인간의 동작을 촬영하기 위해 바이브 트래커 6 개와 베이스 스테이션 2 개를 사용하여 트래커 데이터를 수신받는다. 소프트웨어 개발 도구로는 OpenVR SDK를 활용하였다. 본 장에서 수행하는 실험의 경우 전신 동작 복원 모델과 손 동작 생성 모델 학습 시 사용하지 않은 테스트 모션 데이터셋을 활용하여 실험을 진행하였다. 실험의 경우 크게 3가지로 수행하였는데, 먼저 오토인코더에서 복원된 동작의 정확성를 측정하기 위해 테스트 모션 데이터셋의 동작과 생성된 동작간의 차이를 측정하고, 복원된 동작에 적절한 손 동작을 생성하는 실험도 수행하였다. 두 번째 실험으로 오토인코더에서 복원된 동작의 딜레이를 조절하여 실시간 트래커 데이터와 비교하고, 오토인코더에서 최소 1 에서 최대 10 프

레임 딜레이를 가지고 있는 동작 생성이 가능함을 확인한다. 마지막으로 생성된 동작의 발 미끄러짐 을 방지하기 위해 IK 솔버를 사용하여 보정한 동작과 보정하지 않은 동작에 대한 정량적 비교를 하고 차이를 확인한다.

4.1 손 동작을 포함한 전신 동작 생성

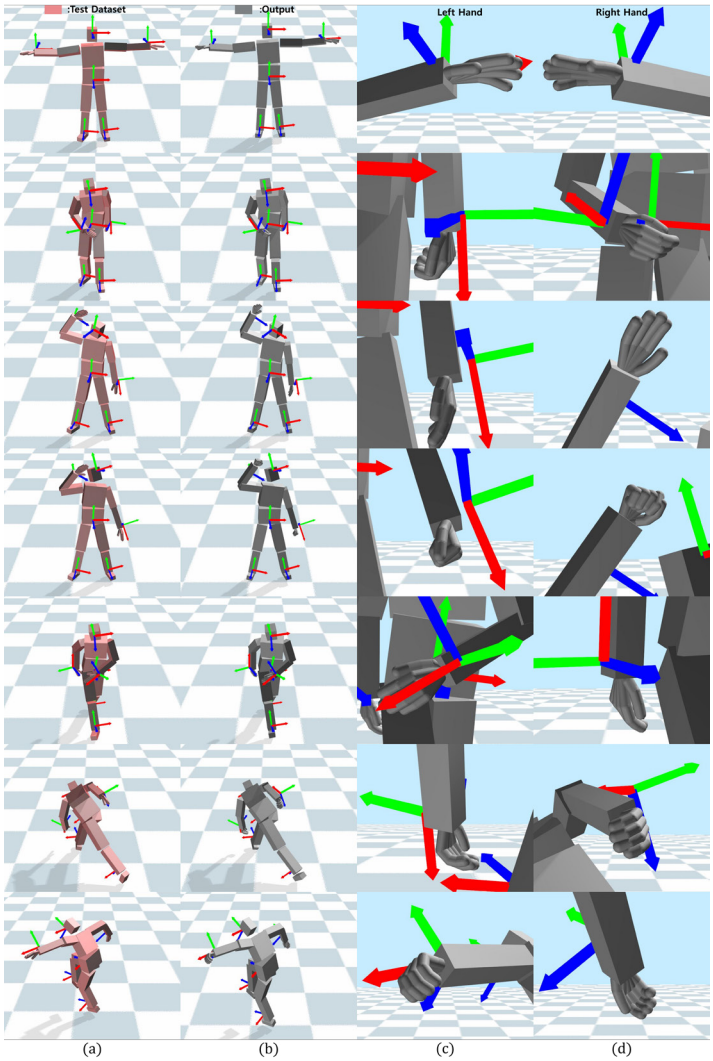


Figure 3: This is the result of motion reconstruction performed using the method proposed in this paper. (a) Ground-truth test motion (not used during training). (b) The resulting motion generated using the Autoencoder and FCN networks proposed in the paper. (c), (d) A close-up view of hand motion generated by an FCN for each corresponding frame.

본 실험에서는 전신 동작 모델인 오토인코더에서 복원된 전신 동작을 확인하고 이에 대한 정량적 평가를 수행한다. 또한 손 동작 생성모델에서 오토인코더에서 복원된 전신 동작을 활용하여 이에 적절한 손 동작이 생성되는 것을 확인한다. Figure 3 은 전신 동작 복원 모델과 손 동작 모델의 결과이다. Figure 3 의 각 행은 같은 시점에 캡쳐된 결과이다. 왼쪽 첫 열의 빨간색의 캐릭터의 동작은 테스트 모션 데이터셋의 동작이고, 회색 캐릭터의 동작은 본 논문에서 제안한 방법을 사용하여 생성한 결과이다. 오른쪽 두 열의 경우 각 시점에서 손 동작 모델에 의해 생성된 결과를 나타낸다. 생성된 결과를 확인해 보았을 때 실시간으로 변화하는 트래커 데이터에 대응하여 캐릭터의 동작이 자연스럽게 생성되는 것을 확인하였다. 또한 빨간색 캐릭터의 경우 손 동작이 생성되지 않아 같은 동작을 반복하고 있는 반면 회색 캐릭터의 손 동작은 전신 동작이 변화할때마다 다양한 동작이 나타났고 이를 통해 손 동작 모델에 의해 캐릭터의 전신 동작에 따라 자연스러운 손 동작이 생성되는 것을 확인하였다. 또한 오토인코더에서 복원된 동작의 정확도를 측정하기 위해 본 실험에서는 전신 동작에 대한 정량평가를 수행하였다. 이를 위해 복원된 동작에 대해 MPJPE(Mean Per Joint Position Error), MPJRE(Mean Per Joint Rotation Error) 그리고 지터링(Jittering) 과 같은 수치를 계산하여 복원된 동작에 정확성을 측정하였다. MPJPE 는 테스트 모션 데이터셋의 캐릭터 동작과 전신 동작 모델의 결과 동작간의 각 관절의 위치 차이를 측정하는 지표이고, MPJRE는 동작간의 관전 회전 값의 차이를 측정하는 지표이다. 지터링 지표는 캐릭터 동작의 불규칙한 변화나 흔들림을 나타낸다. Table 1은 본 논문에서 제안한 오토인코더의 방법과 캐릭터의 양 팔과 발 관절을 엔드이펙터로 지정하고 IK 솔버를 활용하여 동작을 복원한 결과를 비교한 정량평가이다. 각 지표 아래에 있는 화살표가 아래방향을 가리키면 낮을수록 정확한 동작을 오른쪽 방향을 가리키면 GT(Ground Truth)인 모션 데이터셋과 값이 비슷할수록 좋은 결과를 의미한다. Table 1 의 수치를 확인해본 결과 복원된 동작과 테스트 모션 데이터셋의 동작과의 차이를 나타내는 MPJPE와 MPJRE에서 적은 오차를 가지는 것을 확인하였고 이를 통해 오토인코더를 통해 정확한 전신 동작 복원이 가능하다는 것을 확인하였다. 지터링 지표에서도 IK 솔버을 사용한 방법보다 오토인코더의 결과가 모션 데이터셋과 근접한 값을 가지는 것을 확인하였다. 여러 지표들을 사용하여 비교해본 결과 본 논문에서 제안한 오토인코더의 결과가 더욱 정확한 동작 생성하는 것을 확인하였다.

	MPJPE(cm) ↓	MPJRE(degree) ↓	Jittering($10^2 m/s^3$) GT 0.2566 →
IK	8.2206	21.3987	0.2650
AE	4.8057	20.5673	0.2613

Table 1: Quantitative comparison between motion reconstruction using an IK solver based method and the method proposed in the paper.

4.2 동작 복원 딜레이 조절

본 실험에서는 오토인코더에서 복원된 동작들 즉 서로 다른 딜레이를 가지는 동작에 대한 실험을 수행한다. 본 실험에서 딜레이는 프레임 단위로 조절되며, 오토인코더는 최소 1 프레

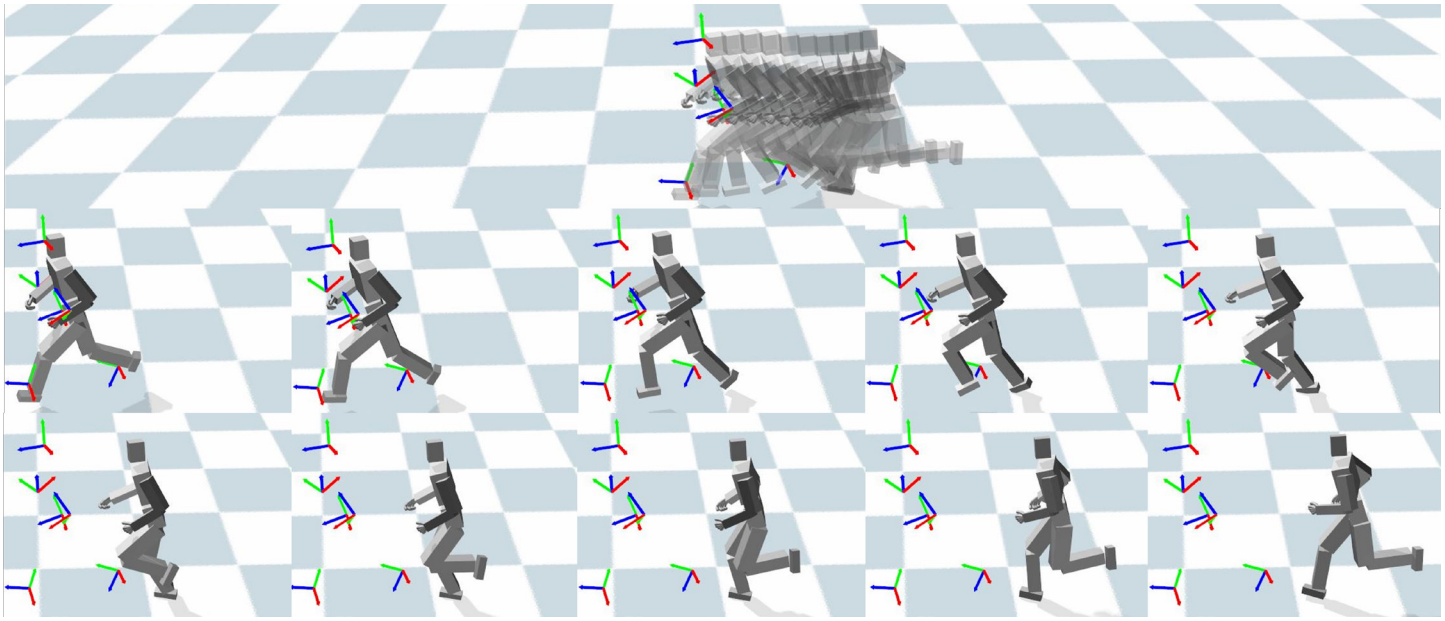


Figure 4: The result of motion generation with various frame delays using real-time tracker data.

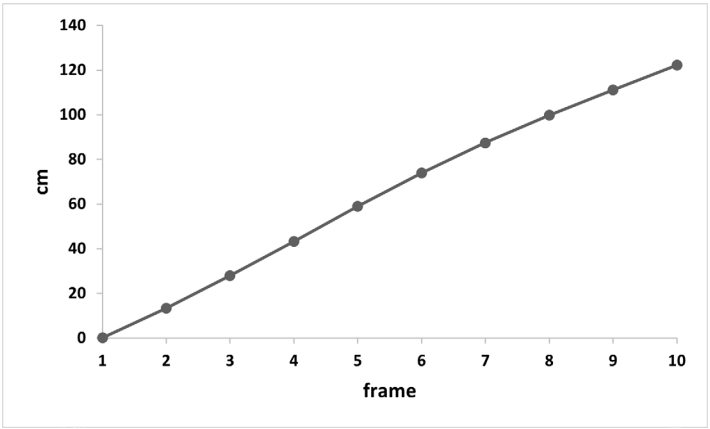


Figure 5: The average distance between trackers and character joints.

임부터 10 프레임의 딜레이를 가지는 동작을 복원할 수 있다. 실험을 위해 테스트 모션 데이터셋에서 임의의 시점인 t 의 트래커 데이터를 추출하고 이를 오토인코더에 입력으로 사용하여 동작을 복원한다. Figure 4 은 테스트 모션 데이터셋에서 임의의 t 시점에 복원된 동작들의 대한 결과이다. 그림에 있는 6개의 좌표계는 테스트 모션 데이터셋에서 t 시점의 트래커 데이터 즉 실시간 트래커의 좌표를 나타내고 t 시점에서 트래커는 약 3.8m/s로 움직이고 있다. 첫 번째 행의 그림은 t 시점에 오토인코더를 통해 복원된 모든 동작들을 시각화한 결과이다. 첫 번째 행의 그림을 통해 오토인코더에서 복원하는 동작들이 각 프레임 딜레이에 따라 연속적이고 안정적인 것을 확인하였다. 두 번째와 세 번째 행의 그림은 각 프레임 딜레이에 생성된 동작과 t 시점의 트래커 데이터를 시각화한 결과이다. 두 번째 행에서 세 번째 행으로, 왼쪽에서 오른쪽열로 갈수록 큰 딜레이 프레임을 가지는 동작의 결과이다. 결과에서 확인할 수 있듯이 왼쪽 열보다 오른쪽 열에 복원된 동작들이 실시간 트래커 데이터와의 거리가 멀어지는 것을 확인할 수 있다. 이를 수치적으로 측정하기 위해 Figure 4에서 딜레이 프레임에 따라 트래커와 캐릭터에 상응하는 관절의 평균 거리를 계산하였다. Figure 5 는 평균 거리를 그래프화한 결과이다. 평균 거리 그래프를 확인해본 결과 딜레이 프레임이 늘어날수록 트래커와 캐릭터간의 거리가 늘어나는 것을 확인할 수 있다. 또한 1 프레임 딜레이를 가지는 동작의 경우 실시간 트래커와의 평균 거리가 1cm 미만으로 트래커의 움직임에 즉시 반응하는 것을 확인하였다. 이를 통해 오토인코더를 통해 여러 딜레이 프레임을 가지는 동작을 복원할 수 있음을 확인하였고 최소 딜레이 프레임을 활용하여 실시간 동작 복원이 가능함을 확인하였다.

4.3 IK 솔버를 활용한 발 미끄러짐 보정

본 실험에서는 발 미끄러짐 현상을 해결하기 위해 IK 솔버에 관련된 실험을 수행한다. IK 솔버를 활용한 보정 성능을 측정하기 위해 IK 솔버를 사용한 동작과 사용하지 않은 동작에 대한 비교를 수행한다. 본 실험에서는 정확한 발 미끄러짐 현상의 측정을 위해 수식 (7) 의 $foot\ skating1$[22] 과 $foot\ skating2$ 수식을 활용하여 각 동작에 대한 발 미끄러짐 을 정량적으로 측정하였다. $foot\ skating1$ 에서 v 는 캐릭터 발의 속력을 나타내고, H 는 캐릭터 발의 최대 높이를 h 는 현재 캐릭터의 발 높이를 나타낸다. 본 실험에서는 최대 발 높이 H 를 바닥으로부터 25cm 로 설정하여 계산하였다. $foot\ skating2$ 에서 c 는 바닥과 캐릭터 발의 접촉을 나타내는 변수로 접촉시 1 비접촉시 0 으로 설정된다.

$$
\begin{aligned}
foot\ skating1 &= v \cdot (2 - 2^{h/H}), \\
foot\ skating2 &= c \cdot v.
\end{aligned} \tag{7}
$$

cm/frame	FS1 ↓	FS2 ↓
motion dataset	0.8118	0.2649
AE without IK	1.0677	0.5132
AE with IK	0.8043	0.2610

Table 2: The average foot skating in motion capture data, autoencoder without IK solver and autoencoder with IK solver.

Table 2은 테스트 모션 데이터셋, 오토인코더를 활용하여 복원한 동작 그리고 오토인코더의 결과에 IK 솔버 를 사용한 동작에 대한 발 미끄러짐 현상을 측정한 결과이다. 두 가지 수식을 통한 결과를 비교해보았을 때 IK 솔버 를 사용하여 보정한 동작이 보정이 적용되지 않은 동작에 비해 발 미끄러짐 현상이 개선되는 것을 확인할 수 있었다. 또한 모션 데이터셋의 동작에서보다 IK 솔버 를 사용한 동작에서 발 미끄러짐 이 덜 발생하는 것을 Table 2 를 통해 확인할 수 있었다. 이를 통해 IK 솔버 를 활용한 동작 보정이 발 미끄러짐 현상을 효과적으로 개선하는 것을 확인하였다.

5 결론

본 논문에서는 트래커를 활용한 딥러닝 기반 동작 복원 방법을 제안한다. 인간의 동작을 촬영하기 위해 6개의 바이브 트래커를 사용하고, 학습된 오토인코더를 통해 정확하고 다양한 딜레이를 가지는 캐릭터 전신 동작을 복원한다. 여러 프레임의 전신 동작을 복원함에 따라 딜레이를 프레임 단위로 조절 가능하며, 실시간 동작 복원까지 가능하다. 복원된 전신 동작에 자연스러운 손 동작을 생성하기 위해 전신 동작을 입력으로 활용하여 손 동작을 생성하는 학습된 FCN 을 사용한다. 이를 통해 트래커 데이터에 따라 적절한 전신 동작을 복원하고 복원된 동작에 적절한 손 동작 생성이 가능함을 실험을 통해 확인하였다. 이후 생성된 전신 동작에서 발 미끄러짐 현상을 개선하기 위해 IK 솔버를 사용하여 캐릭터 하체 자세에 대한 보정을 수행하고 이를 통해 발 미끄러짐 현상이 감소하는 것을 실험을 통해 확인하였다. IK 솔버를 활용한 보정을 통해 정확한 캐릭터 발의 제어가 가능하며 생성된 동작의 역학적 안정성을 보장한다. 제안된 방법은 트래커를 활용한 딥러닝 기반 동작 복원 방법으로 모션 캡쳐 시스템의 현실적인 활용 가능성을 강조하며, 다양한 응용분야에서 동작 생성과 보정 기술의 중요성을 높이는 데 기여할 것으로 기대된다.

하지만 제안된 방법에도 한계점이 존재한다. 손 동작 생성시 사용자의 의도를 파악하여 생성하는 것이 아니라 전신 동작에 적절한 손 동작을 생성하여 정확한 제어가 불가능하다. 이러한 문제를 해결하기 위해 향후 연구에서는 사용자의 의도를 파악하여 그에 적합한 손 동작을 생성함으로써 더욱 세밀한 제어를 목표로 한다. 또한 물리 시뮬레이션을 접목하여 다른 사물과 상호작용이 가능한 시스템 개발에 대한 연구를 계획하고 있다.

감사의 글

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2020-0-01373, 인공지능대학원지원(한양대학교))과 2023년도 패러블엔터테인먼트의 지원을 받아 수행된 연구임 (1425179536, 메타버스 콘텐츠 제작을 위한 올인원 프로그램 개발)

References

[1] T. Sweeney, "Foundational principles & technologies for the metaverse," in ACM SIGGRAPH 2019 Talks, SIGGRAPH '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[2] A. Chatzitofis, G. Albanis, N. Zioulis, and S. Thermos, "A low-cost realtime motion capture system," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21421–21426, 2022.

[3] J. Kim, D. Kang, Y. Lee, and T. Kwon, "Real-time interactive animation system for low-priced motion capture sensors," Journal of the Korea Computer Graphics Society, vol. 28, no. 2, pp. 29–41, 2022.

[4] B. Van Hooren, N. Pecasse, K. Meijer, and J. M. N. Essers, "The accuracy of markerless motion capture combined with computer vision techniques for measuring running kinematics," Scandinavian Journal of Medicine & Science in Sports, vol. 33, no. 6, pp. 966–978, 2023.

[5] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," Sports Medicine - Open, 2018/06/05.

[6] A. Shafaei and J. Little, "Real-time human motion capture with multiple depth cameras," 2016 13th Conference on Computer and Robot Vision (CRV), pp. 24–31, 2016.

[7] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "Monocular 3d human pose estimation in the wild using improved cnn supervision,"

in 3D Vision (3DV), 2017 Fifth International Conference on, IEEE, 2017.

[8] Y. Zou, J. Yang, D. Ceylan, J. Zhang, F. Perazzi, and J.-B. Huang, "Reducing footskate in human motion reconstruction with ground contact constraints," in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 448–457, 2020.

[9] Y. Lu, H. Yu, W. Ni, and L. Song, "3d real-time human reconstruction with a single rgbd camera," Applied Intelligence, vol. 53, p. 8735–8745, aug 2022.

[10] P. Caserman, A. Garcia-Agundez, and S. Goebel, "A survey of full-body motion reconstruction in immersive virtual reality applications," IEEE Transactions on Visualization and Computer Graphics, vol. 26, pp. 3089–3108, 2020.

[11] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time," ACM Trans. Graph., vol. 37, dec 2018.

[12] X. Yi, Y. Zhou, and F. Xu, "Transpose: Real-time 3d human translation and pose estimation with six inertial sensors," ACM Trans. Graph., vol. 40, jul 2021.

[13] M. Kim and S. Lee, "Fusion poser: 3d human pose estimation using sparse imus and head trackers in real time," Sensors, vol. 22, no. 13, 2022.

[14] D. Yang, D. Kim, and S.-H. Lee, "Lobstr: Real-time lower-body pose prediction from sparse upper-body tracking signals," Computer Graphics Forum, vol. 40, 2021.

[15] K. Ahuja, E. Ofek, M. Gonzalez-Franco, C. Holz, and A. D. Wilson, "Coolmoves: User motion accentuation in virtual reality," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 5, jun 2021.

[16] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu, "Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors," 2022.

[17] A. Winkler, J. Won, and Y. Ye, "Questsim: Human motion tracking from sparse sensors with simulated avatars," SA '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[18] F. G. Harvey, M. Yurick, D. Nowrouzezahrai, and C. Pal, "Robust motion in-betweening," vol. 39, no. 4, 2020.

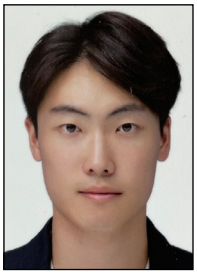
[19] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," 2020.

[20] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," 2021.

[21] L. Kovar, J. Schreiner, and M. Gleicher, "Footskate cleanup for motion capture editing," in Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '02, (New York, NY, USA), p. 97–104, Association for Computing Machinery, 2002.

[22] H. Zhang, S. Starke, T. Komura, and J. Saito, "Mode-adaptive neural networks for quadruped motion control," ACM Trans. Graph., vol. 37, jul 2018.

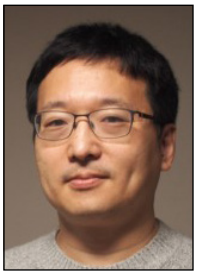
〈저자소개〉



김현석
- 2016-2022 아주대학교 산업공학 학사
- 2022-2024 한양대학교 컴퓨터소프트웨어학 석사과정
- https://orcid.org/0000-0002-5109-7397

강경원
- 2016-2020 중부대학교 정보보호학 수료
- 2020-2022 한양대학교(ERICA) 소프트웨어학 학사
- 2022-현재 한양대학교 컴퓨터소프트웨어학 석사과정
- https://orcid.org/0009-0004-7328-1760

박강래
- 2011-2017 성결대학교 멀티미디어공학 학사
- 2017-현재 한양대학교 컴퓨터소프트웨어학 석박사 통합과정
- https://orcid.org/0009-0002-3366-7453



권태수
- 1996-2000 서울대학교 전기컴퓨터공학부 학사
- 2000-2002 서울대학교 전기컴퓨터공학부 석사
- 2002-2007 한국과학기술원 전산학전공 박사
- https://orcid.org/0000-0002-9253-2156