Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.34, NO.3, Jun. 2024 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2024.34.3.493

메타버스 보안 강화를 위한 동작 기반 사용자 인증*

박 성 규,[1†] 류 권 상[2‡]
[1]숭실대학교 (대학원생), [2]공주대학교 (교수)

Motion-Based User Authentication for Enhanced Metaverse Security*

Seonggyu Park,[1†] Gwonsang Ryu[2‡]
[1]Soongsil University (Graduate student), [2]Kongju National University (Professor)

요 약

본 논문에서는 메타버스 환경 내에서의 사용자 지속 인증 문제에 대해 다룬다. 최근 메타버스는 개인의 상호작용, 엔터테인먼트, 교육 및 비즈니스 분야에서 중요한 역할을 하고 있으며, 특히 사용자 신원 확인과 관련된 취약점이 주요한 문제로 인식되고 있다. 본 연구는 자세 추정 모델로 메타버스 환경의 캐릭터 움직임을 추출하고 분석하여 사용자의 신원을 확인하는 새로운 방법을 제안한다. 이 방법은 영상 데이터만을 이용하여 사용자를 인증하기 때문에 제한적인 환경에서도 활용할 수 있으며, 다양한 실험을 통해 캐릭터의 움직임이 사용자 식별에 어떻게 기여할 수 있는지를 분석한다. 또한, 이 접근 방식의 다른 디지털 플랫폼으로의 확장 가능성을 탐구한다. 이러한 연구는 메타버스 환경 내에서의 보안 강화와 사용자 신원 확인 방식의 혁신에 중요한 기여할 것으로 기대된다.

ABSTRACT

This paper addresses the issue of continuous user authentication within the metaverse environment. Recently, the metaverse plays a vital role in personal interaction, entertainment, education, and business, bringing forth significant security concerns. Particularly, vulnerabilities related to user identity verification have emerged as a major issue. This research proposes a novel method to verify identities by analyzing users' character movements in the metaverse through a pose estimation model. This method uses only video data for authentication, allowing flexibility in limited environments, and investigates how character movements contribute to user identification through various experiments. Furthermore, it explores the potential for extending this approach to other digital platforms. This research is expected to significantly contribute to enhancing security and innovating user identity verification methods in the metaverse environment.

Keywords: Metaverse, Continuous Authentication, Deep Learning

I. 서 론

본 논문은 최근 기술 발전의 전면에 등장한 메타버스 환경 내의 보안 문제를 깊이 있게 다룬다. 메타버스는 현실 세계를 모사하거나 상상력을 기반으로 한 가상 세계를 제공함으로써, 개인의 상호작용, 엔터테인먼트, 교육 및 비즈니스 분야에서 중요한 역할을 하고 있다[1]. 그러나 이러한 가상 환경의 확장은 여러 보안상의 위협을 동반한다. 특히, 사용자 신원 확인과 관련된 취약점이 주요한 문제로 부상하고 있다[2]. 메타버스 서비스를 이용하는 과정에서 사용자들은 대화와 활동을 통해 자신도 모르게 사생활을 드러내기 쉽다. 이러한 상호작용은 사용자들이 자신의 취향, 선호, 쇼핑 목록과 같은 사적인 정보를

Received(03. 18. 2024), Modified(05. 20. 2024), Accepted(05. 24. 2024)
* 본 연구는 2024년도 정부(과학기술정보통신 부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (RS-2023-00229400, 안전한 메타버스 환경을 위한 사용자 인증 및 프라이버시 보호 기술개발)
† 주저자, parksk99@soongsil.ac.kr
‡ 교신저자, gsryu@kongju.ac.kr(Corresponding author)

무의식적으로 드러낼 수 있으며 심지어 서비스의 몰입도를 높이기 위해 결제 정보, 신상 정보 등을 직접 서비스에 제공하기도 한다. 이는 메타버스가 다양하고 깊은 종류의 개인정보를 포함하고 있음을 의미하며, 이 정보는 때때로 공개되기도 한다. 만약 악의적인 공격자가 정상적인 사용자의 계정을 해킹한다면, 이러한 민감한 개인정보를 탈취할 수 있다. 더욱이, 최근에는 생성형 AI의 발달[3]로 특정 인물의 얼굴이나 목소리를 모방[4]하는 것이 가능해져, 계정 보안이 더욱 중요해졌다. 따라서 현재 접속 중인 사용자가 실제 계정의 주인이 맞는지 확인하는 기술은 메타버스 환경에서 안전을 보장하기 위해 필수적이며 이를 위해, 본 연구에서는 지속 인증을 사용한 새로운 접근 방식을 제시하고자 한다.

지속 인증[5]은 사용자의 신체적, 행동적 특성을 연속적으로 모니터링하고 분석함으로써 신원을 확인하는 기술이다. 이러한 접근 방식은 전통적인 단일 시점 인증 방식의 한계를 극복하고, 보안을 강화하는 데에 중요한 역할을 할 수 있다. 지속 인증에는 얼굴이나 홍채를 관찰해 인증하는 생체 지속 인증 방법이 있고 터치 스크린 조작 방법[21]이나 키보드 입력 패턴[22], 걸음걸이[6], 시선[7] 등의 행동을 모델링하는 행동 지속 인증 방법이 있다. 하지만 메타버스 환경은 얼굴이나 홍채 등의 사람의 생체 정보가 드러나지 않기 때문에 생체 인증을 사용하기 어렵다. 또한 터치 스크린이나 키보드를 사용하지 않고 시선을 파악하기 위해서는 고가의 장비가 필요하기 때문에 물리적 환경 기반 행동 인증을 사용하기 어렵다. 따라서 본 연구에서는 걸음걸이 기반 지속 인증을 확장하여 메타버스 환경에 드러나는 사용자의 자세를 추정하여 시간에 따른 행동 변화를 모델링함으로써 사용자를 식별하고 인증하는 방법을 탐구한다.

본 연구에서 제안한 인증 시스템은 VR 게임, 특히 'Beat Saber'에서 VR 기기를 통해 사용자의 행위를 반영한 아바타의 동작만을 분석함으로써 사용자를 인증한다. 이러한 접근 방식은 고가의 생체 인식 하드웨어 센서나 복잡한 설치 과정 없이도 사용자의 독특한 동작 패턴을 활용한다. 따라서, 기존의 생체 인식 시스템에 비해 상당한 센서 비용 절감을 기대할 수 있다.

이를 위해 본 연구는 다양한 실험을 통해 캐릭터의 움직임이 사용자 식별에 신뢰성 있는 인증 요소가 될 수 있는지를 분석한다. 또한, 이러한 접근 방식이 다른 플랫폼에도 적용될 수 있는지를 검토하여, 기술의 범용성과 확장성에 대해 논의한다. 이러한 연구는 메타버스 환경 내에서의 보안을 강화하고, 사용자 신원 확인 방식을 혁신하는 데 중요한 기여를 할 것으로 기대된다.

본 논문은 다음과 같이 구성된다. 2장에서 메타버스 보안, 지속 인증 등 연구의 토대가 되는 관련 연구를 언급하고 3장에서 아바타의 행위를 포함하는 동영상을 수집하는 방법을 정리한다. 4장에서 동영상으로부터 피처를 추출하는 방법을 설명하고 피처를 이용해 분류기와 인증기를 학습하는 과정을 말한다. 5장에서 제안 방법의 기대효과와 한계점을 명시하고 6장에서 결론과 향후 연구를 서술하며 마무리한다.

II. 관련 연구

2.1 메타버스 보안 및 사생활 분석 및 보호 연구

메타버스와 같은 사회적 소통 서비스에는 다양한 보안과 사생활 위협이 존재한다. 거기에 더해 메타버스만의 특성으로 인해 그 문제가 더 가중된다. Yan Huang 등의 연구[2]에 따르면 메타버스는 소셜화, 몰입적 상호작용, 실세계 구축, 확장성의 네 가지 핵심 특성이 있다. 이러한 특성들이 결합하여 더욱 복잡하고 중요한 보안 및 사생활 문제를 야기한다. 개인정보 유출, 감청, 비인가 접근, 피싱 등과 같은 다양한 위협이 존재하며, 몰입도가 높은 사회적 사이버 공간 구축을 위해 더 많은 개인적 요소가 포함됨에 따라 개인 정보 유출 위험이 커진다. 사용자는 소셜 서비스를 즐기기 위해 자신의 프로필이나 행동 데이터 등의 사용자 정보를 제공[8,9]할 수 있으며, 이는 공격자에게 유용한 정보가 될 수 있다. 메타버스에서는 다양한 회사들이 가상의 인생을 구축하는데, 이 과정에서 사용자의 개인정보가 유출될 위험이 있다. 또한 의료 서비스와 치료를 제공하는 건강 관리 메타버스 플랫폼에서 의사로 가장한 악의적인 사용자가 특정 사용자를 속여 위험한 약물을 처방[8]할 수 있으며 특히, 딥러닝 기술을 이용한 가짜 목소리나 외모 생성으로 인해 사용자의 금융 정보 등이 탈취될 수도 있다. 또한 타인의 행세를 하며 성추행 등의 사이버 범죄[10]가 발생할 여지가 있다.

선행 연구에는 메타버스의 보안과 사생활 취약점에 대해 분석[2,8]하였지만, 플랫폼이 준수해야 할 가이드라인만 제안할 뿐 실제적인 방어 방법이 전무하다. 이러한 문제들은 메타버스의 확장성과 사회화

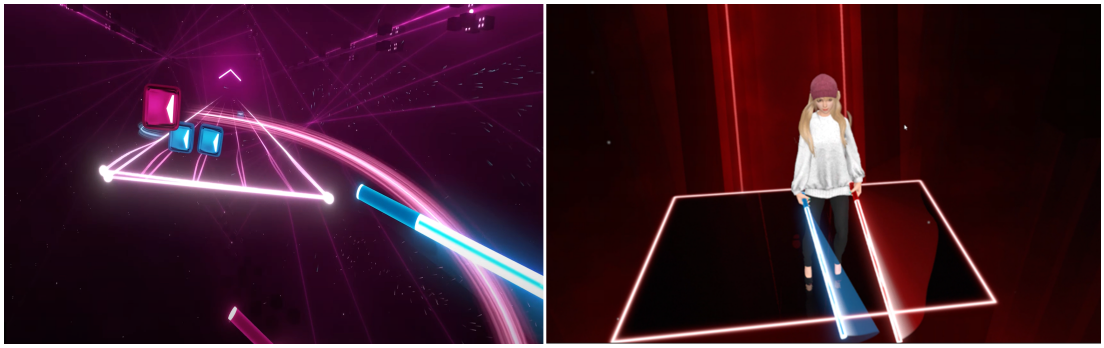

Fig. 1. (Left): Beat Saber gameplay screen. (Right): Screen recorded from a third-person perspective for posture extraction.

특성에 따라 더욱 심각해질 수 있으므로 새로운 형태의 보안 및 사생활 보호 방법이 필요하다.

2.2 지속 인증 연구

지속 인증은 사용자의 신체적, 행동적 특성을 연속적으로 모니터링하고 분석하여 신원을 확인하는 기술이다. 이는 전통적인 단일 시점 인증 방식의 한계를 극복할 수 있으며 특히 메타버스와 같은 디지털 환경에서 중요한 역할을 한다. 지속 인증은 사용자의 행동 양식, 상호작용 방식, 신체적 특성 등을 실시간으로 분석하여 인증 프로세스를 수행한다. 이러한 지속 인증의 개념을 바탕으로, Gattulli 등의 연구[11]는 터치스크린 상호작용을 통한 인증 방법에 초점을 맞췄다. 이 연구에 따르면, 스마트폰의 기본 탐색 작업에서 나오는 촉각 데이터를 분석하여 사용자를 인증하는 시스템을 개발하였다. 이 시스템은 하나의 클래스를 구분하는 Support Vector Machine(SVM) 모델을 학습시켜 98.9%의 정확도와 99.4%의 F1 점수를 달성했다. 이러한 결과는 터치스크린 상호작용을 이용한 사용자 인증이 효과적일 수 있음을 보여주며 메타버스 환경에서의 사용자의 행동 분석을 통한 인증 방식의 구현 가능성을 제시한다.

2.3 행위 인증 연구

본 연구는 실제 환경의 행위 인증 연구를 메타버스 환경으로 확장한 것이다. 실제 환경의 행위를 기준으로 사람을 인증하는 연구는 다양하고 높은 인증 성능을 보인다. 보행 기반 인증 연구인 BRITTANY[12]는 CNN을 사용하여 사람들의 걸음걸이를 분석함으로써 개인을 식별한다. 이 과정에서 LIDAR 데이터를 개인의 보행을 나타내는 이미지로 변환한 뒤 CNN이 이를 처리하여 그들의 독특한 걸음걸이 패턴을 기반으로 개인을 인증한다. 이 방법은 사람의 행동을 신뢰할 수 있는 생체 인증 수단으로 사용할 수 있는 잠재력을 강조한다.

2.4 제안 방법 관련 연구

2.4.1 Human Pose Estimation

인간 자세 추정(Human Pose Estimation, HPE)은 컴퓨터 비전 분야에서 중요한 연구 주제로, 광범위한 응용 분야에서 활용되고 있다. 최근의 발전은 주로 딥러닝 기반 접근 방식에서 이루어졌다.

Shih-En Wei 외 3명[13]은 다단계 처리를 통해 점진적으로 관절 위치의 정확도를 높이는 방식을 채택하는 CPM(Convolutional Pose Machines)을 고안했다. 각 단계는 이전 단계에서 생성된 신뢰도 맵을 활용하여 신체 부위의 위치를 점차적으로 정제한다. CPM은 복잡한 자세와 가려짐이 있는 장면에서도 강인한 성능을 보여주며, 다단계 접근 방식을 통해 정확도를 향상시키고 신체 부위 간의 공간적 관계를 효과적으로 모델링한다.

Jingdong Wang외 11명[14]은 다중 해상도 하위 네트워크를 병렬로 연결하고 반복적인 다중 스케일 융합을 수행하여 높은 해상도 표현을 학습하는 HRNet을 개발했다. 이로 인해 더 정확한 키 포인트 히트맵 예측이 가능하며 HRNet은 해상도 감소 과정에서 정보 손실을 최소화하면서 다양한 해상도의 특징을 효과적으로 결합한다. HRNet의 변형으로는 Lite-HRNet이 있으며, 채널 간 및 해상도 간 정보

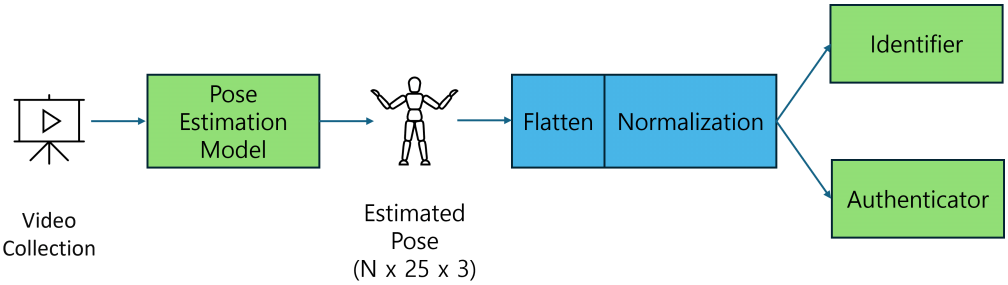

Fig. 2. Identification and Authentication System Architecture

교환을 위해 조건부 채널 가중치 블록을 설계했다.

Zhe Cao외 4명[15]은 Convolutional Pose Machines을 기반으로 한 OpenPose를 개발했다. 히트맵을 통해 키 포인트 좌표를 예측하고, Part Affinity Fields(PAFs)를 사용하여 각 사람에게 키 포인트를 연결하며 PAFs는 2D 벡터 필드의 집합으로, 사지의 위치와 방향을 나타내는 벡터 맵으로 구성된다. OpenPose는 하향식 접근법을 사용하여 처리 속도를 향상시켰다.

이러한 모델들은 HPE 분야에서 혁신적인 발전을 이끌었으며, 각각의 방식은 다양한 응용 분야에서 중요한 역할을 하고 있다. OpenPose의 경우 실시간 처리 및 다중 인물 추적에 효과적이며, HRNet은 높은 해상도 및 정확도를 제공한다. CPM은 다단계 처리를 통해 복잡한 자세를 효과적으로 분석할 수 있다. 이러한 기술들은 향후 HPE 연구 및 응용 개발에 중요한 기반을 제공한다.

2.4.2 Transformer

트랜스포머 모델[16]은 자연어 처리(NLP) 분야에서 혁신적인 심층 학습 아키텍처로, 주로 언어 번역, 텍스트 분류 및 텍스트 생성과 같은 작업에 사용된다. 이 모델은 기존 순환 신경망(RNN)이나 합성곱 신경망(CNN)과 달리 어텐션 메커니즘만[17]을 사용하여 입력 시퀀스 내의 요소 간의 의존성을 모델링한다. 언어 번역, 텍스트 분류 및 텍스트 생성과 같은 다양한 작업에서 최신 결과를 달성하는 데 사용되었다. 이 아키텍처는 여러 변형을 통해 다양한 작업과 문제에 적응할 수 있는 유연성을 가지고 있다.

특히 트랜스포머의 인코더 구조는 입력 시퀀스(sequence)를 처리하고 입력 정보를 요약하는 숨겨진 표현을 생성한다. 이러한 매핑(mapping)은 BERT[18]에서 사용하는 방식으로 시퀀스의 클래스 분류 문제에 탁월한 성능을 보인다. 특히 인코더-디코더 아키텍처에서 인코더는 입력 시퀀스 전체에 대해 관심을 가지고 인코더 구조는 입력 시퀀스를 고정 길이의 표현으로 인코딩하며, 분류기나 회귀 분석기에 사용되는 입력으로 활용될 수 있다.

III. 데이터 수집

행동 기반 지속 인증을 위한 메타버스의 행위에 대한 오픈 데이터는 현재 전무하다. 따라서 'Beat Saber'[19]에서 아바타의 행동적 특징을 뽑아내기 위한 환경을 직접 구축하고 데이터를 수집하였다. 'Beat Saber'는 가상현실(VR) 환경에서 진행되는 리듬 기반의 게임으로, 사용자는 음악에 맞춰 나타나는 블록을 가상의 검으로 자르는 활동을 수행한다. 이 게임을 선택한 이유는 이것이 Fig. 1.의 왼쪽처럼 3차원 메타버스 환경을 제공하여 사용자의 움직임을 자세히 관찰할 수 있으며, 명확한 목표와 과제로 인해 사용자가 게임에 몰입함으로써 무의식적인 행동 양식이 드러나기 때문이다. 연구를 위해 진행된 동영상 수집 과정에서는 'Beat Saber'의 기본 설정인 1인칭 시점을 Fig. 1. 오른쪽에 보이는 것처럼 Camera2 모드를 활용하여 3인칭 시점으로 변경하였으며, Custom Avatar 모드를 통해 캐릭터의 형태를 생성하여 사용자의 움직임을 명확히 포착하였다. 총 6명의 참가자로부터 각각 세 개의 노래에 대해 3회씩 플레이하는 영상을 수집하였으며, 모든 영상은 30fps의 프레임 속도로 일관되게 녹화하였고 평균 129.94초 최소 86.52초 최대 185.41초의 길이를 가지고 있다. 이렇게 수집된 동영상 데이터는 메타버스 환경에서 사용자의 행동 패턴 분석에 있어 핵심적인 기초 자료로 활용될 것이다.

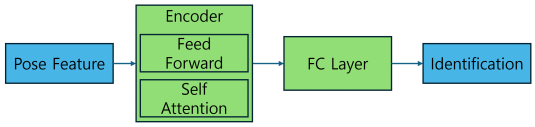

Fig. 3. Identification System Architecture

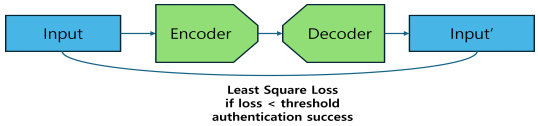

Fig. 4. Authentication System Architecture

IV. 제안 방법

4.1 특징 추출

본 연구에서는 'Beat Saber'라는 게임을 이용하여 사용자의 행동 데이터를 수집하고 이를 분석하는 방법을 개발하였다. 이 게임은 사용자가 가상 공간에서 음악에 맞춰 블록을 자르는 활동을 하는 VR 게임으로, 이를 통해 얻은 동영상 데이터는 사용자의 움직임과 행동 패턴을 파악하는 데 사용된다. OpenPose[15]라는 오픈소스 자세 추정 모델을 사용하여 각 프레임에서 캐릭터의 자세 정보를 추출하며 (T, 25, 3) 형태의 출력을 제공한다. 여기서 T는 동영상의 프레임 수를 나타내며 총 25개의 관절에 대한 x 좌표, y 좌표, confidence 점수가 도출된다.

이러한 자세 데이터는 정규화 과정을 거쳐 최종적인 특징으로 사용된다. 정규화 과정은 사용자의 초기 위치가 인증 프로세스에 미치는 영향을 최소화하여, 모델이 움직임의 본질적인 특성을 더 잘 파악하고 구분할 수 있게 한다. 정규화는 사용자마다 다른 초기 위치를 표준화하여, 모델이 사용자의 움직임 변화를 더 정확하게 구분할 수 있도록 돕는다. 이 과정은 메타버스 환경에서 사용자 인증의 정확성을 높이는 데 핵심적인 역할을 하며 정규화 방법은 5장에서 자세히 설명한다.

4.2 분류기

사용자를 식별하는 모델로 트랜스포머의 인코더 구조를 사용한다. 트랜스포머 인코더는 디코더에 비해 분류, 회귀 등의 문제에 더 좋은 성능을 내기 때문에 본 구조를 사용하였다. 구조는 Fig. 3.와 같으며 자세 추정 모델을 통해 얻은 사용자의 자세 데이터를 트랜스포머의 인코더에 입력으로 넣고 이를 요약하여 식별에 필요한 핵심 정보를 추출한다. 요약된 핵심 정보를 Fully Connected(FC) Layer에 넣어 사용자를 식별한다. 이 과정은 사용자의 식별을 위한 정확하고 효율적인 방법을 제공하여 행동을 정확하게 분석할 수 있다. 분류기의 인코더에 사용한 파라미터는 아래와 같다.

- encoder layer: 8
- attention head: 8
- model dimension: 64
- learning rate: 1e-3

4.3 인증기

메타버스 환경 내 사용자 인증을 위해 오토인코더(auto-encoder) 구조를 채택하였다. 오토인코더는 특정 사용자의 행동 패턴을 학습하고, 이를 기반으로 본인 여부를 판단하는 데 적합한 구조를 가진다. 이는 오토인코더가 입력 데이터의 중요한 특징을 효율적으로 학습하고 추출할 수 있기 때문이다. 또한, 오토인코더는 레이블이 필요 없는 비지도 학습 모델로, 이를 통해 사용자의 정상적인 행동 데이터만으로도 효과적인 인증 모델을 구축할 수 있다. 구조는 Fig. 4.를 따르며 학습 단계에서 원본과 복원 값 사이의 Least Sqaure Error(LSE) 값을 줄이도록 학습한다. 그리고 인증 단계에서 임의의 인증 단위를 넣었을 때 복원 값과 입력값의 LSE 값이 임계값보다 작으면 본인이라고 판단하고 임계값보다 크면 타인이라고 판단한다.

V. 실험 및 검증

3장에서 제안한 방법을 검증하기 위해 4장에서 다양한 방법으로 실험하고 분석하며 의의를 파악한다.

5.1 피처 선택 실험

Fig. 2.의 자세 추출 모델을 통과하면 25개의 관절에 대해 각각 x, y 좌표와 confidence 점수가 도출된다. x와 y값은 평면상의 관절의 위치를 나타내는 좌표값이고 confidence 점수는 자세 추정 모델이 해당 관절의 위치를 예측한 정확도, 즉 실제로 관

절이 그 위치에 존재할 확률을 수치화한 것으로 0과 1사이의 값을 가진다. Confidence 점수를 피처에 포함할 경우 손이 등 뒤로 가거나 무릎이 손에 가려지는 등 신체부위가 가려져 관절의 좌표가 부정확하게 측정되는 경우를 모델이 고려하여 학습할 수 있기 때문에 성능이 향상될 것이라고 예상하고 confidence 점수의 포함 여부를 달리하여 실험을 수행했다. 일반적인 성능을 측정하기 위해 두 개의 window size로 실험했으며 Table 1.에 정리한 바와 같이 두 window size 모두 confidence 점수 유무에 따라 일관된 식별 정확도를 보여준다. 정확도는 아래 수식과 같이 정의된다.

$$(Accuracy) = \frac{(Correct\,\mathrm{Predictions})}{(Total\,\mathrm{Predictions})} \quad (1)$$

실험 결과, confidence 점수가 있을 때 성능이 향상되는 것을 확인할 수 있었다. 이것은 신체 부위가 가려진 상황에서 confidence 점수가 낮게 측정되면 분류 모델이 이것을 고려하여 사용자를 분류할 수 있기 때문이다. 따라서 이후 실험에서는 25개 좌표에 대한 (x, y, confidence 점수)를 선택하여 총 75차원의 피처를 사용한다. 이러한 접근 방식은 관절 위치의 불확실성을 효과적으로 관리하고, 모델의 인식 능력을 최대화하는 데 중요한 역할을 한다.

5.2 인증 단위길이 결정 실험

지속 인증은 인증 단위의 길이(window size)가 중요하다. 그 이유는 인증 성능과 시스템의 가용성에 큰 영향을 미치기 때문이다. 인증 단위는 길이가 짧으면 가용성이 늘어나지만 인증 성능이 떨어지고, 단위의 길이가 길면 인증 성능이 높아지지만 가용성이 떨어지는 trade-off 관계를 형성한다. 따라서 가용성과 인증 성능을 모두 고려해 적절한 단위길이를 설정해야 한다. Fig. 5.는 단위길이에 따른 분류기의 정확도를 그래프로 그린 것이다. 가로축은 인증 단위

Table 1. User Identification Accuracy with and without Confidence

window size	w/ confidence	w/o confidence
300	79.81%	77.41%
100	80.96%	76.48%

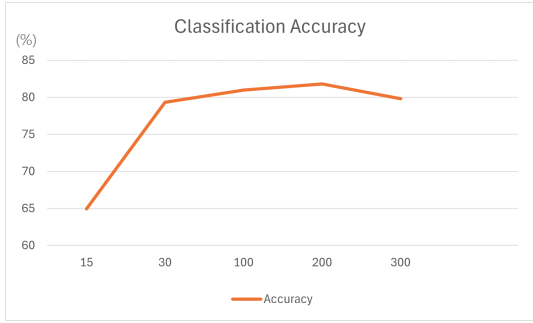

Fig. 5. User Identification Accuracy Based on Authentication Unit Length

길이이고 세로축은 분류기의 정확도를 나타낸다. 길이가 200일 경우를 제외하고 길이가 짧아질수록 낮은 성능을 보인다. 이는 길이가 짧을 경우 더 적은 사용자의 행동적 특징이 담기기 때문이다.

5.3 피처 정규화 방법 실험

자세 추정 모델을 통해 추출한 피처는 사용자가 처음 시작하는 위치에 따라 다른 값을 가지기 때문에 분류기가 사용자의 행동이 아닌 단순히 서 있는 위치를 보고 사용자를 분류하는 문제가 존재한다. 따라서 이 문제를 해결하기 위해 피처를 추출한 후 정규화를 수행한다. 인증 단위의 첫 번째 프레임을 기준으로 정규화하는 시간 정규화, 각 프레임에 나타나는 자세의 중심점을 기준으로 정규화하는 관절 정규화, 앞선 두 방법으로 도출한 값을 이어 붙이는 concat 정규화로 세 가지의 정규화 방법을 실험하여 선택한다. Fig. 6.는 정규화 방법을 설명한 그림이고 그 실험 결과를 Fig. 7.에 정리하였다. 시간 정규화 방법에 비해 관절 정규화나 concat 정규화 방법은 인증 단위의 길이가 짧아져도 성능이 유지된다. 그 이유는 시간 정규화 방법은 첫 프레임으로부터 얼마나 움직였는지를 표현하기 때문에 단위길이가 짧을 경우 그 움직임이 하나의 인증 단위에 모두 담기기 어렵기 때문이다. 또한 같은 단위길이일 때 관절 정규화가 더 높은 성능을 보인다. 이것은 관절 정규화는 중심 관절을 기준으로 얼마나 멀리 떨어져 있는지를 표현하므로 하나의 토큰 안에 사람의 자세 정보가 담겨 보다 정확히 행동을 모델링할 수 있기 때문이다.

관절 정규화와 concat 정규화 방법은 Fig. 7.에서 보다시피 대부분의 경우에 비슷하거나 관절 정규화가 더 높은 정확도를 보인다. 이는 관절 정규화 방

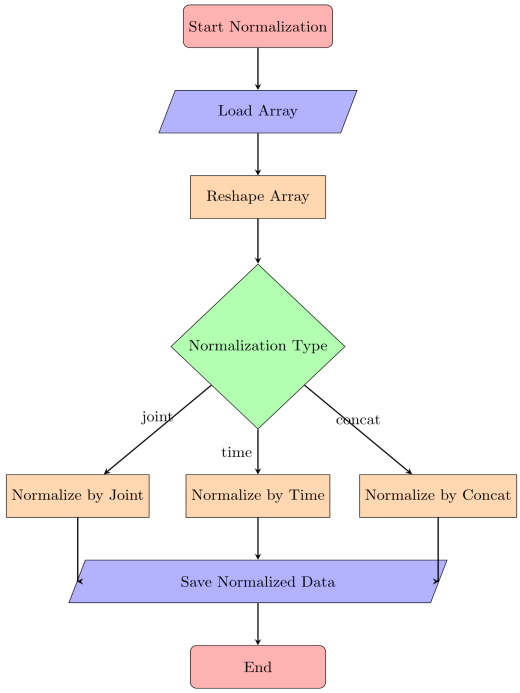

Fig. 6. Normalization Algorithm

법이 독자적으로 정규화를 잘 수행할 수 있는 능력이 있기 때문이라고 판단하고 계산 비용을 줄이기 위해 다른 실험에서도 관절 정규화 방법을 사용한다.

5.4 노래 교차 검증

제안 방법이 사용하는 인증 기준의 확장 가능성과 전이성을 파악하기 위해 학습과 검증에 같은 노래를 사용하는 것이 아닌 다른 노래를 사용해 실험하였다. Table 2.에 보이는 바와 같이 1번 노래를 이용해 학습한 뒤 3번 노래로 검증한 결과 31.10%의 낮은 식별 정확도를 보였지만 1번과 2번 노래를 같이 학습한 뒤 3번 노래로 검증한 결과 두 배 이상 향상된 66.92%의 식별 정확도를 달성했다. 이를 통해 만약 다양한 상황에서 수집한 영상으로 학습한다면 모델이 사용자의 행동 양식을 풍부하게 인지할 수 있어 범용

Table 2. User Identification Accuracy When Validated with a Different Song from the Learned Song

	Accuracy(%)
train-1, test-3	31.10
train-1,2 test-3	66.92

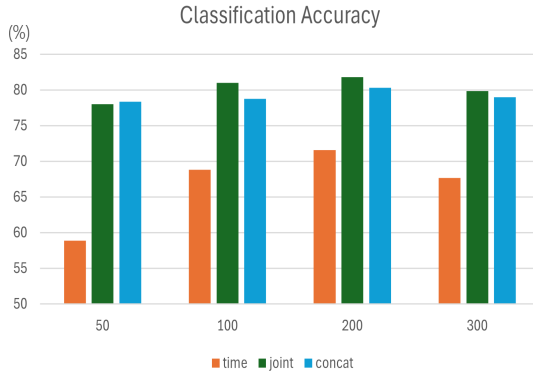

Fig. 7. User Identification Accuracy According to Each Normalization Method

적인 사용자 인증 모델을 구축할 수 있을 것이다.

5.5 인증기 실험

사용자를 인증하기 위해 Fig. 4.의 인증기를 사용한다. 특정한 사용자의 데이터만으로 원본과 유사한 출력을 만드는 방식으로 학습시킨 뒤 인증 과정에서 임의의 데이터가 들어왔을 때 입력과 출력의 차이가 임계값보다 적을 때 본인이라고 간주하고 그렇지 않을 때 타인이라고 판단하여 사용자 인증 Area Under the Curve(AUC)값을 계산하였다. AUC는 ROC 곡선의 아랫 부분 면적을 의미하며 ROC 곡선은 임계값에 따른 False Positive Rate(FPR)과 True Positive Rate(TPR)의 변화를 기준으로 그려지기 때문에 임계값에 따른 모델의 성능을 나타내고 면적이 1에 가까울수록 좋은 성능을 의미한다. 여기서 FPR과 TPR은 아래 수식으로 정의되며, 정상 인증 시도를 positive라고 하고 비정상 인증 시도를 negative라고 하며 이를 옳게 구분한 경우를 true 틀린 경우를 false라고 한다.

$$FPR = \frac{FP}{TN + FP} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

그 결과 Fig. 8.에 ROC 곡선을 작성하였고 AUC 값을 계산하였다. 그 결과 AUC 값은 0.783으로 나타났다.

VI. 고 찰

6.1 기대효과

이 논문의 제안하는 지속 인증 방법은 메타버스 보안 강화에 중요한 기여를 할 것으로 기대된다. 첫째, 메타버스 계정 탈취로 인해 발생할 수 있는 다양한 위협을 해결하는 효과적인 방법을 제시한다. 계정 탈취는 사용자의 개인 정보 유출, 금융적 손실, 그리고 사생활 침해 등 다양한 위험을 수반[2,8,10]한다. 본 논문에서 제안하는 지속 인증 방식은 이러한 위험을 상당 부분 완화할 수 있으며, 메타버스 사용자의 안전을 보장하는 데 중요한 역할을 할 것이다.

둘째, 본 연구는 네트워크 사용량의 감소를 통한 효율성 향상을 기대할 수 있다. Vivek Nair 외 6명의 연구[20]에서는 센서 정보를 이용한 사용자 식별에 232개의 피처를 사용했다는 점을 고려할 때, 본 논문의 제안 방법이 사용하는 75개의 피처는 이에 비해 훨씬 적은 양이다. 이는 동일한 인증 과정을 수행함에 있어 필요한 네트워크 트래픽이 세 배 이상 감소함을 의미한다. 따라서, 본 논문의 방법은 피지컬 환경에서의 행동 기반 지속 인증보다 네트워크 자원을 훨씬 적게 소모하며, 이는 네트워크 트래픽의 부담을 줄이고 전체 시스템의 효율성을 증가시킬 것이다.

마지막으로, 본 연구는 비용 절감의 잠재력을 가진다. 생체 인증 시스템은 고가의 센서 및 복잡한 설치 과정을 요구하는 반면, 본 논문에서 제안하는 방법은 추가적인 하드웨어 없이 VR 게임에서 얻은 데이터를 활용한다. 이는 특히 비용에 민감한 조직이나 개인 사용자에게 매력적인 대안이 될 수 있으며, 인증 시스템의 구축 및 유지 관리 비용을 상당히 절감할 수 있을 것이다.

6.2 한계점

제안 방법은 가상현실(VR) 게임 'Beat Saber'의 플레이 동영상을 분석하여 사용자의 무의식적인 자세 변화를 추출하고, 이를 통해 신원을 확인하는 방식을 포함한다. 이러한 접근 방식은 기존 인증 방법에 비해 메타버스에서의 인증 문제 해결을 위한 새로운 가능성을 열어준다. 그러나, 본 연구는 몇 가지 한계점을 내포하고 있다.

연구에 사용된 데이터의 양과 다양성에 관한 한계

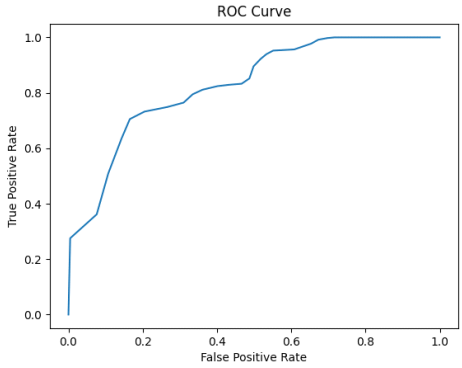

Fig. 8. User Authentication ROC Curve

는 본 연구의 결과가 실제 메타버스 환경의 복잡성과 다양성을 전면적으로 대변하고 있다고 보기 어렵게 만든다. 본 연구는 특정 VR 게임인 'Beat Saber'에서 획득한 동영상 데이터에 기반하여 사용자 인증 방법을 개발하고 검증하였다. 이 게임은 특정한 동작과 리듬 감각을 요구하지만, 메타버스 환경 전체에서 사용자들이 수행할 수 있는 행동의 범위와 다양성을 완전히 포괄하지는 못한다. 따라서, 본 연구의 방법론이 다양한 유형의 활동, 상황, 그리고 사용자 행동 패턴을 포괄적으로 인증할 수 있을지에 대한 질문이 남는다. 더욱 다양한 데이터와 시나리오에 대한 실험이 필요하며, 이는 메타버스의 다양한 환경과 상황에서도 효과적으로 작동할 수 있는 인증 시스템의 개발로 이어질 수 있다. 다양한 가상 환경에서의 사용자 행동과 상호작용의 복잡성을 고려할 때, 본 연구의 방법론이 다른 VR 게임이나 메타버스 활동에 쉽게 적용될 수 있을지에 대한 추가적인 연구와 실험이 요구된다.

또한 본 연구는 메타버스 환경에서 이용자의 움직임을 수집하는 과정에서 프라이버시 위협이 발생한다. 사용자의 움직임을 2차적으로 가공해 어떤 행동을 했는지 알 수 있기 때문이다. 따라서 메타버스 서비스에 실제로 적용할 때 데이터 수집 시 데이터 익명화나 암호화 방법을 사용해 프라이버시를 보호해야 한다.

VII. 결 론

사용자 인증의 신뢰성과 효율성을 높이기 위해 사용자의 행동이 담긴 캐릭터의 움직임을 분석해 인증하는 새로운 접근법을 제시한다. 기존 지속인증은 추가적 장비가 필요하거나 비용이 많이 발생하는 문제

가 있기 때문에 메타버스 환경에서 활용하기 어렵다. 따라서 본 논문에서는 플레이 동영상만으로 사용자의 행동을 분석해 사용자를 인증하였다. Beat Saber 플레이 동영상을 직접 수집한 뒤 자세 추정 모델로 피처를 추출하였다. 추출된 피처를 트랜스포머 모델을 학습시켜 메타버스 환경에서 물리적 센서 데이터가 아닌 아바타의 행위 정보만으로 사용자를 식별하고 인증할 수 있음을 보였다.

향후 연구 계획으로는, 현재의 연구 결과를 바탕으로 추가적인 데이터를 수집하고 다양한 메타버스 환경과 시나리오에서의 적용 가능성을 탐구할 예정이다. 특히, 사용자의 로그 데이터를 활용하여 인증 과정을 더욱 정교화하고, 다양한 사용자 행위 패턴을 포괄할 수 있는 모델을 개발함으로써 메타버스 환경에서의 보안과 개인정보 보호를 향상시킬 것이다.

References

[1] Ning, Huansheng, et al, "A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges.", IEEE Internet of Things Journal, vol.10, no. 16, pp. 14671-14688, Aug. 2023.

[2] Huang, et al, "Security and privacy in metaverse: A comprehensive survey.", Big Data Mining and Analytics, vol. 6, no. 2, pp. 234-247, Jun. 2023.

[3] Bandi, Ajay, et al, "The power of generative ai: A review of requirements, models, input - output formats, evaluation metrics, and challenges.", Future Internet, vol. 15, no. 8, pp. 260, Aug. 2023.

[4] Shoaib, Mohamed R., et al, "Deepfakes, misinformation, and disinformation in the era of frontier ai, generative ai, and large ai models.", International Conference on Computer and Applications, Jun. 2023.

[5] Ryu, Riseul, et al, "A comprehensive survey of context-aware continuous implicit authentication in online learning environments.", IEEE Access, vol. 11, pp. 24561-24573, Mar. 2023.

[6] Rodriguez, et al, "PACE: Providing Authentication through Computational Gait Evaluation with Deep learning.", Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, pp.442-446, Oct. 2023.

[7] Odya, Piotr, et al, "User authentication by eye movement features employing SVM and XGBoost classifiers.", vol. 11, pp. 93341-93353, IEEE Access, Aug. 2023.

[8] Kang, Giluk, et al, "Security and Privacy Requirements for the Metaverse: A Metaverse Applications Perspective.", IEEE Communications Magazine, vol. 62, no. 1, pp. 148-154, Jan. 2023.

[9] X. Chen and K. Michael, "Privacy Issues and Solutions in Social Network Sites", IEEE Technology and Society Magazine, vol. 31, no. 4, pp. 43-53, Dec. 2012.

[10] Tanya Basu, "The metaverse has a groping problem already", https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem/, Dec. 2021

[11] Gattulli, V., et al, "Touch events and human activities for continuous authentication via smartphone", Scientific Reports, vol. 13, no. 1, pp. 10515, Jun. 2023.

[12] Álvarez-Aparicio, et al, "Biometric recognition through gait analysis". Sci Reports, vol. 12, no.1, pp.14530, Aug. 2022.

[13] Wei, Shih-En, et al, "Convolutional pose machines.", Proceedings of the IEEE conference on Computer Vision

and Pattern Recognition, Jun. 2016.
〔14〕 Wang, Jingdong, et al, "Deep high-resolution representation learning for visual recognition.", IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 10, pp. 3349-3364, Oct. 2020.
〔15〕 Cao, Zhe, et al, "Realtime multi-person 2d pose estimation using part affinity fields.", Proceedings of the IEEE conference on computer vision and pattern recognition, Jul. 2017.
〔16〕 Vaswani, Ashish, et al, "Attention is all you need.", Advances in neural information processing systems 30, Dec. 2017.
〔17〕 Bahdanau, Dzmitry, et al, "Neural machine translation by jointly learning to align and translate.", 3rd International Conference on Learning Representations, May. 2015.
〔18〕 Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long and Short Papers), pp. 4171-4186, Jun. 2019.
〔19〕 Beat Games, "Beat Saber", https://beatsaber.com/, Mar. 2024.
〔20〕 Nair, Vivek, et al, "Unique identification of 50,000+ virtual reality users from head & hand motion data.", 32nd USENIX Security Symposium (USENIX Security 23), Aug. 2023.
〔21〕 Leyfer K, Spivak, "A Continuous user authentication by the classification method based on the dynamic touchscreen biometrics." In: Proc of the 24th conf of open innov Assoc (FRUCT) IEEE, Apr. 2019.
〔22〕 Lee H, et al, "Understanding keystroke dynamics for smartphone users authentication and keystroke dynamics on smartphones built-in motion sensors." Security and Communication Networks, Mar. 2018.

〈저 자 소 개〉

박 성 규 (Seonggyu Park) 학생회원
2024년 2월: 숭실대학교 소프트웨어학부 학사
2024년 2월~현재: 숭실대학교 소프트웨어학과 석사과정

류 권 상 (Gwonsang Ryu) 정회원
2016년 2월: 공주대학교 응용수학과 학사
2018년 2월: 공주대학교 융합과학과 석사
2018년 3월~2020년 8월: 공주대학교 융합과학과 박사과정
2020년 9월~2022년 2월: 숭실대학교 융합소프트웨어학과 박사
2022년 3월~2024년 2월: 숭실대학교 사이버보안연구센터 연구교수
2024년 3월~현재: 공주대학교 인공지능학부 조교수