전자통신동향분석, 40권 5호 2025, 52-63

보안 응용을 위한 AI 기반 보행 인식 기술 동향

Trends in AI-Based Gait Recognition Technologies for Security Applications

정다혜 (D.H. Jeong, dhjeong3427@etri.re.kr) 인공지능융합보안연구실 연구원
박소희 (S.H. Park, parksh@etri.re.kr) 인공지능융합보안연구실 책임연구원/기술총괄

ABSTRACT

Artificial intelligence (AI)-driven security systems continue to evolve, with growing demand for contactless biometric identification, making gait recognition a promising solution. Unlike traditional biometric methods, such as face or fingerprint recognition, gait recognition enables long-distance identification without requiring active cooperation from the subject. This paper presents a review of AI-based gait recognition technologies for security applications. It examines the shift from traditional handcrafted approaches to modern deep learning models utilizing data modalities like silhouette, color images (RGB), and skeleton data. The study also explores recent advancements in multimodal fusion, transformer-based architectures, and large-scale dataset development for real-world applications. Additionally, we identify key challenges in the field, including privacy protection, vulnerabilities to adversarial attacks, and the need for computational efficiency in real-time deployment. By analyzing these factors, the paper provides insightinto the current state of gait recognition technology and its potential as a viable biometric solution for security. The findings highlight that although recognition accuracy has significantly improved, addressing security vulnerabilities and deployment constraints remains critical for its practical implementation in real-world security systems.

KEYWORDS AI, Biometric Security, Deep Learning, Gait Recognition, Privacy & Ethical AI

I. 서론

최근 디지털 전환이 가속화되면서 보안 위협의 양상도 갈수록 지능화되고 다양화되고 있다. 이러한 변화로 인해 기존의 생체 인식 기술만으로는 복잡해지는 보안 요구사항을 완전히 충족하기 어려운 상황이 되었다. 이에 따라 새로운 생체 인식 기술에 대한 관심이 높아지고 있으며, 그중에서도 보행 인식 기술이 주목받고 있다. 보행은 개인의 신체 구조와 운동 습관에 따라 고유하게 나타나는 행동적 생

* DOI: https://doi.org/10.22648/ETRI.2025.J.400506
* 이 논문은 2025년도 정부(법무부/과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임[No. RS-2024-00426469, 개인 일상모델 기반 일탈탐지 AI 전자감독 기술개발].

체 정보로서, 원거리 및 비접촉 환경에서도 비교적 안정적인 식별이 가능하다. 따라서 기존의 얼굴 인식, 지문 인식 등 전통적인 생체 인식 수단의 한계를 보완할 수 있는 유망한 대안으로 평가받고 있다[1]. 특히 얼굴이 가려지거나 사용자가 비협조적인 상황에서도 인식할 수 있다는 점은 범죄 수사와 같은 분야에서도 실질적인 이점을 제공한다[2].

초기의 보행 인식 기술은 주로 단순한 외형 정보나 주기적 패턴에 의존했으며, 연구자들이 수동으로 정의한 수치적 특징을 추출하여 분류기에 입력하는 방식으로 구현되었다[3]. 그러나 이러한 전통적인 접근 방식은 조명 변화, 복장 차이, 카메라 시점 변화 등 외부 환경 요인에 취약하여 일관된 인식 성능을 유지하기 어렵다는 한계가 존재한다.

이러한 기술적 한계를 해결하기 위해 딥러닝 기반의 인공지능 기술이 보행 인식 분야에 도입되기 시작했다. 딥러닝 모델은 걸음걸이 데이터로부터 고차원적인 특징을 자동으로 학습할 수 있으며, 시간적 · 공간적 패턴을 통합적으로 분석함으로써 기존 방식의 한계를 크게 개선했다. 특히 딥러닝 기반의 영상 처리 기술, 다중 센서 융합 기법, 3D 자세 추정 등의 발전은 보행 인식 시스템의 정확도와 환경 적응성을 향상시켰다. 이에 따라 보행 인식은 공항, 금융기관, 국방 등 고보안 분야를 넘어 스마트 시티, 헬스케어와 같은 민간 영역으로까지 확장되고 있다.

실제 적용 사례를 살펴보면, 중국에서는 2018년부터 Watrix사의 보행 인식 기술을 공공안전 시스템에 도입하여 얼굴이 가려진 상황에서도 효과적인 인물 식별을 수행하고 있다[4]. 또한, 영국에서는 CCTV 영상을 통한 보행 분석 결과가 법정에서 증거 자료로 인정받는 등[5], 보행 인식 기술의 신뢰성과 실용성이 사회적으로 검증되고 있다. 이러한 흐름은 보행 인식 기술이 연구 단계를 넘어 실질적인 보안 솔루션으로 확장되고 있음을 보여준다.

이에 본고에서는 보행 인식 기술의 발전 과정과 AI 기반 최신 연구 동향을 체계적으로 분석하고, 대표적인 기술 및 모델들의 특징을 상세히 검토한다. 더불어 현재 기술이 당면한 과제와 한계점을 객관적으로 진단하고, 이를 바탕으로 향후 발전 방향을 제시하고자 한다.

II. 연구 동향

보행 인식 기술은 입력 데이터의 표현 방식과 처리 기법에 따라 다양한 연구 방향으로 발전해왔다. 이러한 기술적 진보는 보행 인식의 정확도를 높이는 데 이바지하였으며, 다양한 보안 환경에서도 안정적으로 작동할 수 있는 기반을 마련하게 되었다[6]. 본 장에서는 이러한 발전 과정을 구체적으로 살펴보고, 현재까지의 주요 연구 동향을 분석하고자 한다.

1. 전통적인 보행 인식 기술

초기 보행 인식 기술은 주로 실루엣 기반의 외형적 특징에 의존하는 방식으로 개발되었다. 이 시기의 연구들은 대부분 연구자가 직접 정의한 수치적 특징을 추출하여 활용하는 Handcrafted Feature 기반 접근법을 채택했다. 구체적으로는 실루엣의 가로 폭, 보행 주기, 관절 간 거리 및 각도 등의 특징을 추출한 후, 이를 SVM이나 PCA 등의 전통적인 분류기에 적용하는 방식이었다.

이러한 접근법의 대표적인 예로는 GEI가 있다. GEI는 하나의 보행 주기 내 실루엣 이미지를 평균화하여 정적인 보행 특징을 추출하는 방식으로, 연산량이 적고 구조가 단순하다는 장점을 지닌다[7]. 이와 함께 시간적 움직임 정보를 강조하는 MEI와 MHI도 널리 활용되었으며, 이들은 영상 내 동작의 형태와 방향성을 효과적으로 표현할 수 있었다[3].

하지만 이러한 Handcrafted 방식들은 실제 적용 환경에서 여러 제약사항을 드러냈다. 조명 변화, 복장 차이, 카메라 시점 변화 등 현실적인 환경 조건에 민감하게 반응했으며, 이는 인식 성능의 일관성을 저해했다. 보안 응용 관점에서 볼 때, 이와 같은 방식은 인식 정확도가 낮아 허위 알람 증가, 침입자 식별 실패 등으로 이어져 보안 시스템의 신뢰도를 저하할 수 있다.

이에 따라 다양한 보행 조건과 환경 변화에 강건하게 대응할 수 있는 보행 인식 기술이 요구되었으며, 이와 같은 기술적 필요성에 따라 딥러닝 기반 접근 방식이 주목받기 시작하였다.

2. AI 기반 보행 인식 기술의 필요성

보행 인식은 조명, 복장, 촬영 시점, 배경 등 다양한 환경 요소의 영향을 받는 복잡한 시공간적 패턴을 포함하고 있으며, 이로 인해 전통적인 수작업 기반 특징 추출 방식으로는 높은 인식 정확도를 확보하기 어려웠다[8]. 더욱이 실제 보안 환경에서 발생하는 저화질 영상이나 실루엣 왜곡 상황에서는 정적인 형태 정보만으로는 신뢰할 만한 개인 식별이 불가능했다.

이러한 문제점을 해결하기 위해, 딥러닝을 비롯한 인공지능 기술이 보행 인식 분야에 본격적으로 도입되었다. AI 기반 모델의 가장 큰 장점은 원시 데이터로부터 고차원적이고 추상적인 특징을 자동으로 학습할 수 있다는 점이다. 또한 시간적 변화와 공간적 구조를 동시에 고려하는 통합적 분석이 가능하여, 다양한 보행 조건과 환경 변화에 대해 훨씬 뛰어난 적응력을 보여준다.

또한, 기존의 Handcrafted 방식이 단일 시점이나 제한된 조건에 최적화된 것과 달리, AI 모델은 시점 불변성, 비협조 조건, 실시간 처리 등 보안 응용에서 요구되는 특성을 충족시킬 수 있는 유연한 구조를 갖추고 있다[9].

결과적으로 AI 기반 보행 인식 기술은 정형화되지 않은 실제 환경 상황에서도 높은 식별 정확도와 일반화 성능을 유지할 수 있어, 실제 보안 시스템에서의 실용성과 신뢰성을 동시에 확보할 수 있는 핵심 기술로 부상하고 있다.

3. AI 기반 보행 인식 활용 데이터

3.1 입력 데이터 유형

보행 인식 모델에서 입력 데이터의 유형은 시스템 성능과 활용 환경을 결정하는 주요 요소이다. 입력 형태는 주로 RGB 영상, Skeleton 정보, Silhouette 이미지, Depth Map, Infrared 영상 등으로 구분되며, 각 데이터 유형은 인식 정확도, 환경 적응성, 연산 효율성 측면에서 고유한 특성과 한계를 지닌다[9].

RGB 기반 데이터는 인물의 색상, 질감, 의복 등의 정보를 포함하여 다양한 특징을 추출할 수 있다는 장점이 있으나[10-12], 의류 질감이나 색상에 의한 간섭에 민감할 수 있다[13]. 반면 Silhouette 데이터는 배경 제거를 통해 인물의 외형 정보에 집중할 수 있으며, 대표적으로 GEI 등의 표현 방식이 사용된다. 이러한 실루엣 기반 접근은 구조가 단순하고 연산이 비교적 가볍지만[6,14,15], 의복 변화나 촬영 시점 변화에 쉽게 영향을 받는다는 제약이 있다[16].

Skeleton 기반 데이터는 관절 위치 및 움직임을 중심으로 한 고차원적 시공간 정보를 제공함으로써 자세나 보행 주기에서 발생하는 역동적인 특징을 효과적으로 표현할 수 있다. 특히 관절 간 거리나 각도 기반 표현은 의복 변화나 시점 전환에 비교적 강건한 특성을 보인다[8,9].

이 외에도 Depth 또는 Infrared 기반 보행 데이터는 조도가 낮은 환경이나 야간 감시 환경에서의 활

용 가능성을 제시하고 있으며[17,18], 복합 센서 기반 데이터는 다양한 형태의 입력을 통합적으로 처리하여 인식 성능을 향상하는 데 이바지하고 있다[19]. 최근에는 이러한 다양한 입력 형태들을 결합한 멀티 모달 기반 접근도 활발히 연구되고 있으며, 보안 감시와 같은 실제 환경에서의 적용 가능성을 높이는 방향으로 진화하고 있다.

3.2 공개 데이터셋 및 수집 방식

AI 기반 보행 인식 기술의 성능 향상과 실제 환경에서의 적용 가능성 검증을 위해 다양한 조건을 반영한 대규모 공개 데이터셋 구축이 진행되고 있다. 초기 보행 인식 데이터셋은 통제된 실험실 환경에서 수집되었으나, 점차 야외 및 실제 환경 변수를 포함하는 방향으로 발전해왔다. 표 1은 최근 5년 내 주요 보행 인식 데이터셋을 요약한 결과를 보여준다.

대표적인 공개 데이터셋으로는 OU-ISIR 시리즈[20-28]와 CASIA 시리즈[29-33]가 있다. OU-ISIR 시리즈는 일본 오사카 대학에서 개발된 대규모 데이터셋으로, 다양한 보행 조건과 환경 변화를 반영한다. OU-ISIR Speed는 보행 속도의 영향을 분석할 수 있도록 설계되었으며[20], OU-ISIR LP Bag은 가방을 소지한 상태에서의 변화를 포함한다[24]. 특히 OU-MVLP는 14개 시점에서 10,307명의 보행 영상을 촬영한 멀티뷰 대규모 데이터셋으로, 시점 변화에 강인한 모델 학습에 활용되고 있다[25].

표 1 최근 5년 내 주요 보행 인식 데이터셋

데이터셋	연도	데이터 유형	피험자 수	시점 수	시퀀스 수	수집 환경
OU-MVLP Pose[26]	2020	2D Pose	10,307	14	288,596	Indoor
CASIA-E[33]	2020	Silhouette, Infrared	1,014	26	778,752	Outdoor
ReSGait[34]	2021	Silhouette, 2D Pose	172	1	870	Wild
GREW[35]	2021	Silhouette, 2D/3D Pose, Flow	26,345	882	128,671	Wild
VersatileGait[36]	2021	Silhouette	10,000	44	1,320,000	Unity3D
Gait3D[37]	2022	Silhouette, 2D/3D Pose, 3D Mesh & SMPL	4,000	39	25,309	Wild
OU-MVLP Mesh[27]	2022	3D Mesh & SMPL	10,307	14	288,596	Indoor
CCPG[38]	2023	Silhouette, RGB	200	10	16,566	In/Outdoor
DroneGait[39]	2023	Silhouette, 2D/3D Pose, 3D Mesh & SMPL, Optical Flow	96	30	22,718	Outdoor
GaitLU-1M[40]	2023	Silhouette	1,035,309	1,379	1,035,309	Wild
SUSTech1K[41]	2023	RGB, Silhouette, 3D Point Cloud	1,050	12	25,239	Outdoor
AerialGait[42]	2024	Silhouette, 2D/3D Pose, Parsing	533	10	82,454	Outdoor
CCGR[43]	2024	RGB, Silhouette, Parsing, Pose	970	33	1,580,617	Indoor
Scoliosis1K[44]	2024	Silhouette, Pose	1,050	1	1,493	Indoor
Health & Gait[45]	2025	Silhouette, Semantic Segmentation, Optical Flow, 2D Pose	398	1	1,564	Indoor
GaitDN[46]	2025	2D Pose	1,009	3	3,300	Wild(Night)
OU-MVLP OF[28]	2025	Dense Optical Flow	10,307	14	288,596	Indoor

CASIA 시리즈는 중국과학원 자동화연구소에서 구축한 데이터셋으로, CASIA-B가 대표적인 벤치마크 역할을 하고 있다[30]. 이 데이터셋은 가방, 겉옷 조건과 11개 시점에서 124명의 보행 데이터를 수집했으며, 이후 CASIA-D, E 등 확장 버전을 통해 다양한 변수에 관한 연구 기반을 제공하고 있다[32,33]. 하지만 이러한 초기 데이터셋들은 실제 환경의 복잡성을 충분히 반영하지 못한다는 한계를 보인다. 조명 변화, 배경 다양성, 예기치 못한 행동 등 현실적인 조건들이 제대로 포함되지 않아 실용적 적용에는 제약이 있었다.

이를 보완하기 위해 최근에는 더욱 현실적인 조건을 반영한 데이터셋들이 등장하고 있다. GREW는 26,345명의 피험자와 128,671개의 시퀀스를 포함하는 대규모 야외 데이터셋으로, 연령, 성별, 복장, 시점, 소지품 등 다양한 요인을 반영한다[35]. Gait3D는 대형 마트에서 수집된 25,309개의 시퀀스와 4,000명의 피험자 정보를 기반으로 하며, 실루엣, 2D/3D Skeleton, 3D Mesh, SMPL 등 다양한 형식의 데이터를 포함해 멀티 모달 학습을 지원한다[37]. 이는 단일 데이터 형태의 한계를 극복하고 더 강건한 모델 개발을 가능하게 한다.

또한, 야간 환경에서의 보행 인식 성능을 평가하기 위한 시도도 이루어지고 있다. GaitDN 데이터셋은 낮은 조도 조건에서 RGB와 적외선(IR) 영상을 수집해, 야간 환경에서도 모델의 강건성을 실험할 수 있는 기반을 제공한다[46].

4. AI 기반 보행 인식 모델

4.1 Silhouette 기반 모델

보행 인식 분야에서 Silhouette 기반 접근법은 딥러닝의 발전과 함께 지속적으로 진화해 왔다. 이 방법론은 배경 제거 과정을 통해 보행자의 윤곽 정보를 추출하고, 시간적 변화에 따른 움직임 패턴을 학습하는 것을 핵심으로 한다. 초기 연구들은 주로 GEI와 같은 프레임 평균화 기법을 활용하였으나, 구현의 단순함에도 불구하고 시간적 정보의 손실이라는 근본적 한계를 내포하고 있었다. 딥러닝 기술의 발전과 함께 시계열 데이터 처리에 특화된 모델들이 등장하면서 더 정교한 특징 표현이 가능해졌다.

GaitSet 모델은 기존의 순차적 프레임 처리 방식에서 벗어나 집합(Set) 구조를 도입하였다[6]. 이 접근법은 프레임 순서에 의존하지 않고 보행 특징을 학습함으로써 다양한 시점 변화에 대한 강건성을 확보하였다. 이어서 GaitPart는 인체 구조의 특성을 반영하여 실루엣을 수직 방향으로 분할하는 전략을 채택하였다[47]. 각 신체 부위별로 독립적인 Temporal Feature를 학습함으로써 국소적 움직임 패턴의 차별화된 포착이 가능해졌으며, 이러한 부위별 접근법은 전체적인 인식 성능 향상에 이바지하였다.

GaitGL은 전역-지역 특징 추출 모듈을 통해 Dual-Branch 구조를 구현하였다[48]. Global Feature와 Local Feature를 동시에 활용하는 이 방법론은 Local Temporal Aggregation 기법과 결합하여, 포괄적이면서도 세밀한 보행 표현 학습을 가능하게 하였다.

최근에는 Transformer 구조를 활용한 연구도 진행되고 있다. TransGait는 ViT 구조를 보행 인식에 적용한 대표적 사례로, 패치 기반 이미지 분할과 Self-Attention 메커니즘을 통해 CNN 기반 방법론의 한계를 극복하고자 하였으며, 실험 결과는 복잡한 환경 변화에 대한 적응성을 보여주었다[49].

또한, GAN 기반 접근법도 주목받고 있다. Gait-GAN 모델은 생성적 적대 신경망을 활용해 환경 변화에 불변한 특징 학습을 시도했다[50]. Discriminator와 Generator의 적대적 학습을 통해 조명, 배경, 시점 변화 등의 외부 요인에 강건한 보행 표현을 학습하는 것을 목표로 하였다.

4.2 RGB 기반 모델

RGB 기반 보행 인식 방법론은 원본 비디오 영상을 직접 활용함으로써 보행자의 형태적 특성뿐만 아니라 색상, 질감 등의 풍부한 시각적 정보를 동시에 학습할 수 있다. 이러한 접근법은 별도의 전처리 과정 없이 End-to-End 학습이 가능하다는 점에서 실용적 장점을 제공한다.

초기 RGB 기반 연구에서는 전방 보행 영상 분석 시스템을 통해 관찰 기반 보행 동작 분석의 가능성을 제시하였다[51]. 이 시스템은 보행 특징과 관절 정보 추출을 통해 정상 보행과 이상 보행의 구분이 가능함을 확인하였으며, 재활 및 스포츠 분야로의 적용 가능성을 보여주었다. 이러한 초기 연구를 바탕으로 RGB 영상에서 정교한 특징 추출의 필요성이 인식되면서, Edge 기반 접근법이 주목받기 시작하였다.

GaitEdge는 RGB 영상에서 추출된 Edge 기반 실루엣을 CNN 인코더로 처리하는 End-to-End 프레임워크를 제시하였다[12]. 이 모델은 보행과 무관한 배경 노이즈와 텍스처 정보를 제거함으로써, CASIA-B 및 OU-MVLP 데이터셋에서 우수한 Cross-Domain 성능을 달성하였다. 이후에는 더욱 풍부한 시공간적 정보를 활용하기 위한 연구가 지속되었다.

GaitFormer는 단일 RGB 카메라 입력을 위한 Dual-Stream Spatial-Temporal Vision Transformer 구조를 채택하여 새로운 방향성을 제시하였다[52]. 이 모델은 CNN 모듈의 공간적 특징 추출 능력과 Transformer 모듈의 시간적 관계 모델링을 결합하여, 임상 보행 분석 영역에서의 활용 가능성을 확장하였다. 특히 무릎 굴곡 각도 및 보행 속도 추정과 같은 정량적 분석에서 높은 정확도를 보여주어, RGB 기반 보행 인식의 실용적 응용 범위를 크게 확장시켰다.

4.3 Skeleton 기반 모델

앞서 논의한 Silhouette 및 RGB 기반 접근법과 달리, Skeleton 기반 보행 인식은 인체의 관절 움직임 자체를 구조화된 형태로 학습하는 방법론이다. 이 접근법은 복장이나 배경과 같은 외형적 요소에 상대적으로 덜 민감하여 환경 변화에 대한 안정성을 제공한다는 장점이 있다.

GaitGraph는 RGB 영상에서 추출한 2차원 Skeleton 정보를 그래프 구조로 모델링하여 GCN을 통해 시공간 패턴을 학습하는 방법을 제안하였다[53]. 이 모델은 CASIA-B 및 OU-MVLP Pose 데이터셋에서 기존 실루엣 기반 모델 대비 최대 3배 이상의 성능 향상을 달성하며, 그래프 기반 접근법의 효용성을 입증하였다. 그래프 구조를 활용함으로써 관절 간의 연결 관계와 시간적 변화를 모델링할 수 있었다.

이후 인체의 해부학적 계층 구조를 반영한 Gait Pyramid Transformer 모델도 제안되었다[54]. 이 모델은 공간적 및 시간적 주의 집중 메커니즘을 단계적으로 구성하여 CASIA-B에서 평균 82.6%의 정확도를, GREW 데이터셋에서 52.16%의 성능을 기록하였다. 기존 Skeleton 기반 모델 대비 약 6% 이상의 성능 개선을 보여주어 Transformer 기반 처리 방식의 우수한 일반화 능력을 확인하였다.

또한, SkeletonGait 모델은 관절 좌표를 Gaussian Heatmap 형태의 Skeleton Map으로 변환하여 CNN 또는 Hybrid 네트워크로 처리하는 방식을 제시하였다[55]. 특히 Silhouette과 Skeleton 정보를 통합한 SkeletonGait++는 GREW 데이터셋에서 Rank-1 85% 이상의 정확도를 달성하며 기존 방법론을 크게 뛰어넘었다. 이 접근법은 단일 형태의 데이터를 사용하는 것보다 멀티 모달리티의 방법이 효과적일 수 있다는 점을 보여주었다.

4.4 멀티 모달 기반 모델

지금까지 살펴본 Silhouette과 Skeleton 기반 접근법은 각각 고유한 장점을 보유하고 있으나, 의상이나 배경 변화에 대한 민감성과 외형 정보 부족이라는 개별적 한계를 동시에 극복하기에는 제약이 있다. 이러한 문제를 해결하기 위해 최근에는 두 가지 이상의 정보 유형을 결합하는 멀티 모달 방식이 주목받고 있다. 이 접근법은 서로 다른 모달리티의 상호 보완을 통해 복합적인 보행 표현력을 확보하고, 환경 변화에 강건한 인식 성능을 달성하고자 한다.

BiFusion은 Silhouette과 Skeleton을 함께 처리하여 Skeleton의 구조 정보를 Multi-Scale Gait Graph 네트워크로 추출하고, 이를 실루엣 특징과 융합하여 풍부한 보행 표현을 학습하는 방법을 제시하였다[56]. CASIA-B와 OU-MVLP 데이터셋 실험에서 Rank-1 정확도 92.1%를 기록하며, 두 모달리티의 상호 보완적 특성을 실험적으로 입증하였다. 이 연구는 단일 모달리티의 한계를 극복하기 위한 융합 접근법의 효용성을 보여주었다.

MMGaitFormer은 Transformer 기반 구조를 활용하여 공간 및 시간 차원의 특징 융합을 수행한다[57]. 먼저 Silhouette과 Skeleton을 각각 인코딩한 후, 공간 융합 모듈(SFM)과 시간 융합 모듈(TFM)을 통해 두 모달리티의 상호작용을 반영한다. Fine-grained Body Parts Fusion 및 Cycle Position Embedding 등을 도입하여 특히 의상 변화에서 Rank-1 정확도 94.8%를 달성함으로써 일반화 성능을 입증했다.

TriGait는 세 개의 분기를 통해 Silhouette, Skeleton, RGB 기반 동작 템플릿을 하나의 네트워크 내에서 통합 학습하는 구조를 제안하였다[58]. 각 모달리티 간 Low-Level Feature 정렬 및 Cross-Modal Interaction을 통해 CASIA-B 전체 평균 96.0%, 의상 변화 조건에서 94.3%의 성능을 달성하며 기존 단일 모달리티 방식 대비 뛰어난 식별력을 증명하였다. 이는 세 가지 서로 다른 모달리티의 시너지 효과를 확인하였다.

또한, ZipGait는 Skeleton 데이터를 기반으로 Diffusion 모델을 활용해 실루엣 유사 표현을 재구성하고, 이를 Skeleton과 함께 통합하는 방식을 채택한다[59]. 이를 통해 Cross-Domain 및 Intra-Domain 조건 모두에서 기존 모델들을 능가하는 성능을 보였으며, 모달리티 간 정보 보강 효과를 극대화하였다.

멀티 모달리티 기반 모델들은 단일 모달리티의 단점을 보완할 수 있으며, 특히 환경 변화에 대해 견고함이 요구되는 보안 적용 분야에서 높은 실용 가능성을 보여준다. 그러나 이러한 융합 방식은 모델 복잡도 증가, 모달리티 간 정렬 비용, 학습 및 추론 시 병합 구조 설계의 난이도 등의 기술적 도전을 동반한다. 따라서 향후 실제 보안 환경에 적용하기 위해서는 모달리티 정렬의 효율화, 추론 최적화, 멀티 모달 학습 안정성 확보 등이 중요한 연구 과제로 남아있다.

III. 도전과제

앞서 살펴본 바와 같이 보행 인식 기술은 지속적으로 연구되어왔다(표 2 참고). 그러나 이러한 연구 성과에도 불구하고 실제 보안 환경에서의 실용적 적용을 위해서는 여전히 해결해야 할 문제들이 존재한다. 특히 보안 시스템으로서의 신뢰성과 실용성을 확보하기 위해서는 기술적 한계뿐만 아니라 보안 취약성과 프라이버시 보호 문제에 대한 포괄적 접근이 필요하다.

1. 보안 측면의 문제

1.1 개인정보 및 프라이버시 보호 이슈

보행 패턴은 개인의 고유한 생체 정보로서 얼굴

표 2 주요 보행 인식 AI 모델

모델 구분	주요 모델	연도	입력 데이터 유형	핵심 구조/기법
Silhouette	GaitSet[6]	2019	Silhouette Sequence	Set-Based CNN
	GaitPart[47]	2020	Silhouette Sequence	Body Part Segmentation + Temporal CNN
	GaitGL[48]	2021	Silhouette Sequence	Global-Local Feature Fusion
	TransGait[49]	2022	Silhouette Sequence	Transformer + Self-Attention
RGB	GaitEdge[12]	2022	RGB Video	Edge-Based Encoder
	GaitFormer[52]	2024	RGB Video	Dual-Stream ViT
Skeleton	GaitGraph[53]	2021	2D Skeleton	Graph Convolution Network
	GaitPT[54]	2023	2D Skeleton	Pyramid Transformer
	SkeletonGait[55]	2024	Skeleton Heatmap	CNN + Skeleton Map Fusion
Multi-modall	BiFusion[56]	2021	Silhouette + Skeleton	MSGG + Fusion Block
	MMGaitFormer[57]	2023	Silhouette + Skeleton	Spatial/Temporal Fusion Transformer
	TriGait[58]	2023	Silhouette + Skeleton + RGB	Tri-Branch Fusion Network
	ZipGait[59]	2024	Skeleton + Generated Silhouette	Diffusion + Fusion

이나 지문과 같은 수준의 식별력을 보유한다. 개인의 보폭, 자세, 걸음걸이 특성은 반복적이고 일관되게 유지되며 의도적 변조가 어렵다. 이러한 정보가 무단으로 수집되거나 유출될 경우 개인의 동의 없이 감시, 추적, 행동 분석에 악용될 수 있다.

공공장소의 감시 카메라나 보안 시스템에서 수집되는 보행 정보는 개인의 명시적 동의 없이 저장되고 분석될 수 있어 프라이버시 침해 우려를 초래한다. 보행 특징이 템플릿 형태로 저장되면 비가역적 정보가 되어 재사용이나 공유가 가능하다는 점은 추가적인 문제다.

따라서 보행 인식 기술의 실용화를 위해서는 익명화 기법, 데이터 암호화, 템플릿 비가역화, 민감 정보 제거 등의 프라이버시 보호 메커니즘이 필요하다.

1.2 보행 데이터의 보안 취약성

최신 연구들은 보행 인식 시스템이 적대적 공격에 취약함을 보여준다. 공격자는 실루엣 이미지나 입력 프레임에 미세한 노이즈를 삽입하여 모델의 오인식을 유도할 수 있다. 이는 얼굴이나 지문 인식 시스템에서 관찰되는 적대적 공격과 유사하다.

Universal Adversarial Perturbation과 같은 공격 기법은 소량의 픽셀 변화만으로도 전체 시스템의 오작동을 일으킬 수 있어 보안 시스템으로서의 신뢰성을 저해한다.

따라서 실제 보안 환경에 적용되는 보행 인식 시스템은 공격에 대한 내성을 확보하고 실시간 위협 탐지 및 대응 체계를 갖추어야 한다. 이는 단순한 인식 정확도 향상을 넘어 모델의 신뢰성과 보안성을 보장하는 종합적 접근을 요구한다.

2. 기술적 문제

2.1 다양한 환경 조건에서의 적용 한계

보행 인식 기술은 조명 조건, 촬영 시점, 배경 변화, 복장, 보행 속도 등 다양한 환경 요소에 민감하게 반응한다. 조도가 낮거나 그림자가 많은 야외 환

경에서는 정확한 실루엣 추출이 어려우며, 카메라 시점 변화에 따라 동일인의 보행 패턴도 다르게 관찰될 수 있다.

실험실 환경에서 구축된 데이터셋에서 높은 정확도를 보이는 모델들이 실제 야외 환경에서는 성능이 저하되는 현상이 보고되고 있다. 이러한 실험실과 현실 세계 간의 성능 격차는 기술의 실용적 적용을 제한한다.

이를 극복하기 위한 멀티 모달 데이터 결합, 도메인 적응 기술, 적대적 학습 기반 강건한 모델 설계 등의 연구가 진행되고 있으나, 완전한 일반화 성능을 갖춘 모델의 개발은 여전히 과제로 남아있다.

2.2 최적화 및 실시간 처리 문제

보안 시스템에서는 지연 없는 실시간 처리가 핵심 요구사항이다. 현재의 고성능 보행 인식 모델들은 대부분 딥러닝 기반으로 높은 연산 복잡성을 갖고 있어 대형 GPU나 서버 환경에서만 원활한 추론이 가능하다. Transformer 기반 구조나 Multi Steam CNN, 멀티 모달 모델 등은 메모리와 연산량 요구사항이 높아 경량 기기나 임베디드 환경에서의 구현이 제한된다.

엣지 디바이스에서는 모델 용량, 연산량, 전력 소모 등을 종합적으로 고려해야 하며, 현재 대부분의 최신 모델들은 이러한 제약 조건에서 효율성이 낮다. 따라서 모델 경량화, 지식 증류, 효율적 추론 스케줄링, 하드웨어 가속 최적화 등이 필요하다.

또한, 실시간 환경에서 발생하는 프레임 누락, 네트워크 지연, 조도 변화 등에도 안정적으로 대응할 수 있는 시스템 설계가 요구된다.

IV. 결론

본고는 보안 강화를 위한 핵심 수단으로 주목받고 있는 AI 기반 보행 인식 기술의 발전 과정과 현재 동향을 분석하였다. 전통적인 수작업 기반 접근법의 환경적 취약성을 극복하기 위해 딥러닝 기술이 도입되면서 인식 정확도와 환경 적응성에서 진보를 이루어냈다.

기술적 측면에서는 Silhouette, RGB, Skeleton 등 다양한 입력 형태를 활용한 모델들이 개발되었으며, 멀티 모달 융합 방식과 Transformer 구조의 도입을 통해 복잡한 환경 변화에 대한 강건성이 향상되었다. 대규모 공개 데이터셋 구축과 야외 및 야간 환경을 포함한 다양한 실험 조건의 확대를 통해 실용적 적용 가능성이 검증되고 있다.

그러나 실제 보안 환경에서의 실용화를 위해서는 여전히 과제들이 남아있다. 개인정보 보호와 프라이버시 침해 우려, 적대적 공격에 대한 취약성, 실시간 처리를 위한 계산 효율성 확보 등은 기술적 성능 향상과 함께 해결되어야 할 문제들이다.

향후 연구는 단순한 인식 정확도 개선을 넘어서 실제 적용 환경에서의 신뢰성과 보안성을 보장하는 방향으로 발전해야 한다. 이를 통해 보행 인식 기술은 기존 생체 인식 기술을 보완하는 보안 솔루션으로 자리 잡을 수 있을 것으로 전망된다.

참고문헌

[1] M.S. Nixon et al., “Human identification based on gait,” Springer Science & Business Media, vol. 4, 2010.

[2] J. Han and B. Bhanu, “Individual recognition using gait energy image,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 2, 2005, pp. 316-322.

[3] A.F. Bobick and J.W. Davis, “The recognition of human movement using temporal templates,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 3, 2002, pp. 257-267.

[4] D. Kang, “Chinese ‘gait recognition’ tech IDs people by how they walk,” AP News, 2018. 11. 6. https://apnews.com/article/bf75dd1c26c947b7826d270a16e2658a

[5] I. Bouchrika et al., “On using gait in forensic biometrics,” J. Forensic Sci., vol. 56, no. 4, 2011, pp. 882-889.

[6] H. Chao et al., “Gaitset: Regarding gait as a set for cross-view gait recognition,” in Proc. AAAI Conf. Artif. Intell., (Honolulu, HI, USA), Jan. 2019, pp. 8126-8133.

[7] J.N. Mogan et al., “Gait-ViT: Gait recognition with vision transformer,” Sensors, vol. 22, no. 19, 2022, p. 7362.

[8] A. Sepas-Moghaddam and A. Etemad, “Deep gait recognition: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, 2022, pp. 264-284.

[9] C. Shen et al., “A comprehensive survey on deep gait recognition: Algorithms, datasets, and challenges,” IEEE Trans. Biometrics, Behav. Identity Sci., vol. 7, no. 2, 2024, pp. 270-292.

[10] Z. Zhang et al., “On learning disentangled representations for gait recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, 2022, pp. 345-360.

[11] X. Li et al., “End-to-end model-based gait recognition,” in Proc. Asian Conf. Comput. Vis., (Kyoto, Japan), Nov. 2020, pp. 3-20.

[12] J. Liang et al., “Gaitedge: Beyond plain end-to-end gait recognition for better practicality,” in Proc. Eur. Conf. Comput. Vis., (Tel Aviv, Israel), Nov. 2022, pp. 375-390.

[13] R. Geirhos et al., “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness,” in Proc. Int. Conf. Learn. Represent., (New Orleans, LA, USA), May. 2019, pp. 1-22.

[14] C. Song et al., “GaitNet: An end to-end network for gait based human identification,” Pattern Recognit., vol. 96, 2019.

[15] B. Lin et al., “Gait recognition with multiple-temporal-scale 3D convolutional neural network,” in Proc. ACM Int. Conf. Multimed., (Seattle, WA, USA), Oct. 2020, pp. 3054-3062.

[16] Z. Liu et al., “Studies on silhouette quality and gait recognition,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., (Washington, DC, USA), Jun. 2004.

[17] M.N. Kumar and R.V. Babu, “Human gait recognition using depth camera: A covariance based approach,” in Proc. Indian Conf. Vis. Graph. Image Process., (Mumbai, India), Dec. 2012, pp. 1-6.

[18] S.A.F. Manssor et al., “Gait recognition system in thermal infrared night imaging by using deep convolutional neural networks,” J. Donghua Univ., vol. 38, no. 6, 2021.

[19] Q. Zou et al., “Robust gait recognition by integrating inertial and RGBD sensors,” IEEE Trans. Cybern., vol. 48, no. 4, 2018, pp. 1136-1150.

[20] Y. Makihara et al., “The OU-ISIR gait database comprising the treadmill dataset,” IPSJ Trans. Comput. Vision Appl., vol. 4, 2012, pp. 53-62.

[21] H. Iwama et al., “The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition,” IEEE Trans. Inf. Forensics Security, vol. 7, no. 5, 2012, pp. 1511-1521.

[22] A. Mansur et al., “Gait recognition under speed transition,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., (Columbus, OH, USA), Jun. 2014, pp. 2521-2528.

[23] C. Xu et al., “The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation,” IPSJ Trans. Comput. Vision Appl., vol. 9, no. 1, 2017.

[24] M.Z. Uddin et al., “The OU-ISIR large population gait database with real-life carried object and its performance evaluation,” IPSJ Trans. Comput. Vision Appl., vol. 10, no. 1, 2018.

[25] N. Takemura et al., “Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition,” IPSJ Trans. Comput. Vision Appl., vol. 10, no. 1, 2018.

[26] W. An et al., “Performance evaluation of model-based gait on multiview very large population database with pose sequences,” IEEE Trans. Biometrics, Behav., Identity Sci., vol. 2, no. 4, 2020, pp. 421-430.

[27] X. Li et al., “Multi-view large population gait database with human meshes and its performance evaluation,” IEEE Trans. Biometrics, Behav., Identity Sci., vol. 4, no. 2, 2022, pp. 234-248.

[28] A. Shehata et al., "OUMVLP-OF: Multi-view large population gait database with dense optical flow and its performance evaluation," IEEE Access, vol. 13, 2025, pp. 87100-87111.
[29] L. Wang et al., "Silhouette analysis-based gait recognition for human identification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, 2003, pp. 1505-1518.
[30] S. Yu et al., "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in Proc. Int. Conf. Pattern Recognit., (Hong Kong, China), Aug. 2006, pp. 441-444.
[31] D. Tan et al., "Efficient night gait recognition based on template matching," in Proc. Int. Conf. Pattern Recognit., (Hong Kong, China), Aug. 2006, pp. 1000-1003.
[32] S. Zheng et al., "A cascade fusion scheme for gait and cumulative foot pressure image recognition," Pattern Recognit., vol. 45, no. 10, 2012, pp. 3603-3610.
[33] C. Song et al., "CASIA-E: A large comprehensive dataset for gait recognition," IEEE Trans. Pattern Anal. Mach. Intell., 2022, pp. 1-16.
[34] Z. Mu et al., "ReSGait: The real-scene gait dataset," in Proc. Int. Joint Conf. Biometrics, (Shenzhen, China), Aug. 2021, pp. 1-8.
[35] Z. Zhu et al., "Gait recognition in the wild: A benchmark," in Proc. IEEE/CVF Int. Conf. Comput. Vis., (Montreal, QC, Canada), Oct. 2021, pp. 14789-14799.
[36] P. Zhang et al., "A large-scale synthetic gait dataset towards in-the-wild simulation and comparison study," ACM Trans. Multimedia Comput. Commun. Appl., vol. 19, no. 1, 2023, pp. 1-23.
[37] J. Zheng et al., "Gait recognition in the wild with dense 3D representations and a benchmark," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (New Orleans, LA, USA), Jun. 2022, pp. 20228-20237.
[38] W. Li et al., "An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Vancouver, BC, Canada), Jun. 2023, pp. 13824-13833.
[39] A. Li et al., "Gait recognition with drones: A benchmark," IEEE Trans. Multimedia, vol. 26, 2024, pp. 3530-3540.
[40] C. Fan et al., "Learning gait representation from massive unlabelled walking videos: A benchmark," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 12, 2023, pp. 14920-14937.
[41] C. Shen et al., "LidarGait: Benchmarking 3D gait recognition with point clouds," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Vancouver, BC, Canada), Jun. 2023, pp. 1054-1063.
[42] A. Li et al., "AerialGait: Bridging aerial and ground views for gait recognition," in Proc. ACM Int. Conf. Multimed., (Melbourne, VIC, Australia), Oct. 2024, pp. 1139-1147.
[43] S. Zou et al., "Cross-covariate gait recognition: A benchmark," in Proc. AAAI Conf. Artif. Intell., (Vancouver, BC, Canada), Feb. 2024, pp. 7855-7863.
[44] Z. Zhou et al., "Gait patterns as biomarkers: A video-based approach for classifying scoliosis," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv., (Marrakesh, Morocco), Oct. 2024, pp. 284-294.
[45] J. Zafra-Palma et al., "Health & Gait: A dataset for gait-based analysis," Sci. Data, vol. 12, no. 1, 2025.
[46] H. Li et al., "Exploring gait recognition in wild nighttime scenes," Appl. Sci., vol. 15, no. 1, 2025.
[47] C. Fan et al., "GaitPart: Temporal part-based model for gait recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Seattle, WA, USA), Jun. 2020, pp. 14225-14233.
[48] B. Lin et al., "Gait recognition via effective global-local feature representation and local temporal aggregation," in Proc. IEEE/CVF Int. Conf. Comput. Vis., (Montreal, QC, Canada), Oct. 2021, pp. 14648-14656.
[49] K. Li and S. Meng, "TransGait: Vision transformer based gait recognition network," in Proc. Int. Conf. Image Process., Comput. Vis. Mach. Learn., (Chengdu, China), Nov. 2023, pp. 339-343.
[50] S. Yu et al., "GaitGAN: Invariant gait feature extraction using generative adversarial networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, (Honolulu, HI, USA), Jul. 2017, pp. 30-37.
[51] M. Nieto-Hidalgo et al., "Vision based gait analysis for frontal view gait sequences using RGB camera," in Proc. Int. Conf. Ubiquitous Comput. Ambient Intell., (San Bartolomé de Tirajana, Gran Canaria, Spain), Nov. 2016, pp. 26-37.
[52] J. Li et al., "GaitFormer: Leveraging dual-stream spatial-temporal vision transformer via a single low-cost RGB camera for clinical gait analysis," Knowl.-Based Syst., vol. 295, 2024.
[53] T. Teepe et al., "GaitGraph: Graph convolutional network for skeleton-based gait recognition," in Proc. IEEE Int. Conf. Image Process., (Anchorage, AK, USA), Sep. 2021, pp. 2314-2318.
[54] A. Catruna et al., "GaitPT: Skeletons are all you need for gait recognition," in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., (Istanbul, Turkiye), May 2024, pp. 1-10.
[55] C. Fan et al., "SkeletonGait: Gait recognition using skeleton maps," in Proc. AAAI Conf. Artif. Intell., (Vancouver, BC, Canada),

Feb. 2024, pp. 1662–1669.

[56] Y. Peng et al., “Learning rich features for gait recognition by integrating skeletons and silhouettes,” Multimed. Tools Appl., vol. 83, no. 3, 2024, pp. 7273-7294.

[57] Y. Cui and Y. Kang, “Multi-modal gait recognition via effective spatial-temporal feature fusion,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Vancouver, BC, Canada), Jun. 2023, pp. 17949-17957.

[58] Y. Sun et al., “TriGait: Aligning and fusing skeleton and silhouette gait data via a tri-branch network,” in Proc. IEEE Int. Joint Conf. Biometrics, (Ljubljana, Slovenia), Sep. 2023, pp. 1-9.

[59] F. Min et al., “ZipGait: Bridging skeleton and silhouette with diffusion model for advancing gait recognition,” arXiv preprint, 2024. doi: 10.48550/arXiv.2408.12111