

프라이버시 보호를 위한 3D 포즈 추정 기반 계층적 보안 프레임워크

이동현, 남민석, 오혁진, 김재호*

세종대학교

dojan.sejong@gmail.com, namminseok.sejong@gmail.com,

hyuckjin.sejong@gmail.com, kimjh@sju.ac.kr

A Privacy Preserving Hierarchical Security Framework Using 3D Pose Estimation

Dong Hyun Lee, Min Seok Nam, Hyuck Jin Oh, JaeHo Kim*

Sejong University

요약

본 논문은 인공지능 기반 영상 분석 기술의 발전으로 보안 효율성이 향상되는 동시에 개인정보 및 프라이버시 침해 위협이 증가하는 문제에 주목한다. 이를 해결하기 위해, 3D 포즈 추정과 디지털 트윈을 결합한 다계층 모니터링 기반 지능형 보안 인프라를 제안한다. 제안 시스템은 공간적 맥락과 위협 조건에 따라 분석 수준을 조절하고, 위협 후보에 한해 익명화 된 추적을 수행함으로써 불필요한 개인 정보 추론을 최소화한다. 이를 통해 AI 기반 보안의 효율성과 프라이버시 보호를 구조적으로 양립시키는 보안 프레임워크를 제시한다.

I. 서론

공공 공간이나 다중 이용 시설과 같이 불특정 다수가 출입하는 환경의 보안 시스템은 오랜 기간 CCTV 영상과 경비 인력의 상시 관찰에 기반해 운영되어 왔으며, 공간 내 모든 이용자를 동일한 수준으로 지속적으로 관찰하는 구조를 취해 왔다. 이러한 환경에서는 다수의 영상 스트림을 동시에 관측해야 하므로, 감시 대상이 증가할수록 상황 인지와 대응의 효율성이 관제 인력의 경험과 집중도에 크게 의존하는 한계가 발생한다 [1]. 이러한 제약을 보완하기 위해 최근에는 컴퓨터 비전 및 인공지능 기술을 활용하여 감시 영상을 자동으로 분석하고, 공간 내 객체의 위치, 이동 경로, 특정 보호 대상과의 거리, 행동 패턴 등을 정량적으로 추론함으로써 위험 상황의 조기 인지 가능성을 높이고 관제 부담을 완화하려는 연구들이 활발히 이루어지고 있다 [2,3]

한편, 기존의 다수 AI 기반 영상 분석 시스템은 무인화 및 자동화를 통해 감시 효율을 향상시키는 데 초점을 두고 설계되어 왔으며, 분석 대상의 위험도나 공간적 및 행위적 맥락과 무관하게 모든 보행자의 행동을 동일한 수준으로 처리하는 구조를 유지하는 경우가 많다. 이러한 구조에서는 보안 목적과 직접적인 관련성이 낮은 행동 정보까지 지속적으로 분석될 수 있으며, 분석 대상과 정보 처리 범위 간의 관계에 대한 체계적인 고려가 어렵다. 본 연구는 이러한 한계에 주목하여, 모든 객체를 동일한 정밀도로 분석하는 기존 방식 대신 공간적 맥락과 위험 조건에 따라 인식 수준을 단계적으로 조절하는 LoD(Level of Detail) 기반 인식 구조를 제안한다.

II. 3D 포즈 추정 기반 다단계 보안 프레임워크

본 제안 시스템은 다중 장소 및 다중 시점 CCTV 영상 분석과 디지털 트윈 기술을 결합하여, 보행자의 위치와 공간적 맥락에 따라 분석 정밀도를 동적으로 조절하는 적응형 LoD 기반 지능형 보안 프레임워크를 수행한다. 시

스텝은 모든 보행자를 정밀 분석하는 대신, 위험도와 보호 대상과의 거리에 따라 분석 수준을 LoD 1에서 LoD 3 까지 단계적으로 조절함으로써 보안 효율성을 유지하는 동시에 불필요한 개인 정보 추론을 제한하도록 설계되었다.

사전에 캘리브레이션 된 다중 카메라에서 획득한 2D 영상 프레임을 SelfPose3D[4]의 2D 관절 히트맵 추정과 3D 객체 위치 추정 모듈을 거쳐 장소의 공간 좌표계 상에서 객체의 전역 위치를 추정한다. 첫 번째 단계인 LoD 1에서는 추가적인 인식 추론을 수행하지 않고 추정된 전역 위치 정보만을 활용하여 공간 점유 현황과 이동 흐름을 파악한다. 두 번째 단계인 LoD 2는 객체가 보호 대상과의 상호작용 가능성이 존재하는 근접 영역에 진입했을 때 활성화된다. 이 단계에서 SelfPose3D [4]의 3D 포즈 추정 알고리즘을 가동하여 관절 단위의 정밀한 행동 분석을 수행하되, 얼굴 등 개인 식별 정보는 배제하고 신체 구조와 움직임만을 추상적으로 처리한다.

마지막 단계인 LoD 3는 보호대상의 혼란 가능성이 높은 RoA(Region of Alert)에 진입하여 연속적인 관찰이 필요 할 때 활성화되는 위험 대응 단계이다. 이 단계에서는 객체의 의복 색상과 외형 특징 같은 비식별적 정보를 활용한 익명화 된 재식별(Re-ID)을 통해 임시 고유 ID를 부여하며, 장소별 관측 결과를 기반으로 전역 클러스터링을 위험 맥락의 연속성을 제한적으로 유지한다. 각 LoD 단계별 분석 결과는 Isaac Sim [5] 기반의 디지털 트윈 환경에 실시간으로 동기화된다. 디지털 트윈 상에서 일반 보행자(LoD 1, 2)는 단순한 원통이나 기본 스켈레톤으로 표현되는 반면, 위협 요소를 가진 객체(LoD 3)는 시각적으로 구별되는 경보 스켈레톤으로 시각화 되어 보안 관리자가 직관적으로 상황을 인지하고 대응할 수 있도록 설계되었다.

III. 구현 및 검증

본 연구에서 제안한 LoD 기반 지능형 보안 프레임워크의 동작 가능성을 확인하기 위해, 개념 검증(Proof of Concept, PoC)을 위해 다중 이용자가 보호 대상 주의를 자유롭게 이동하는 가상의 전시관 환경을 구성하고 단계별 동작 사례를 분석하였다. PoC 환경은 다중 장소(장소 1, 장소 2)와 다중 시점(각 장소당 4 개의 영상 입력), 그리고 장소 별 3~4 명의 관람객들이 활동하는 구성을 통해 실제 전시 공간에서의 운용 시나리오를 모사하였다. 관람객들은 전시물 주위에서 이동하거나 대화하는 일반적인 상호작용을 수행하였으며, 시나리오 상에서 위험 후보 객체는 전시물과의 물리적 접촉이 발생하는 상황을 가정한 이후 해당 공간을 이탈하여 다른 장소로 이동하도록 설정하였다.

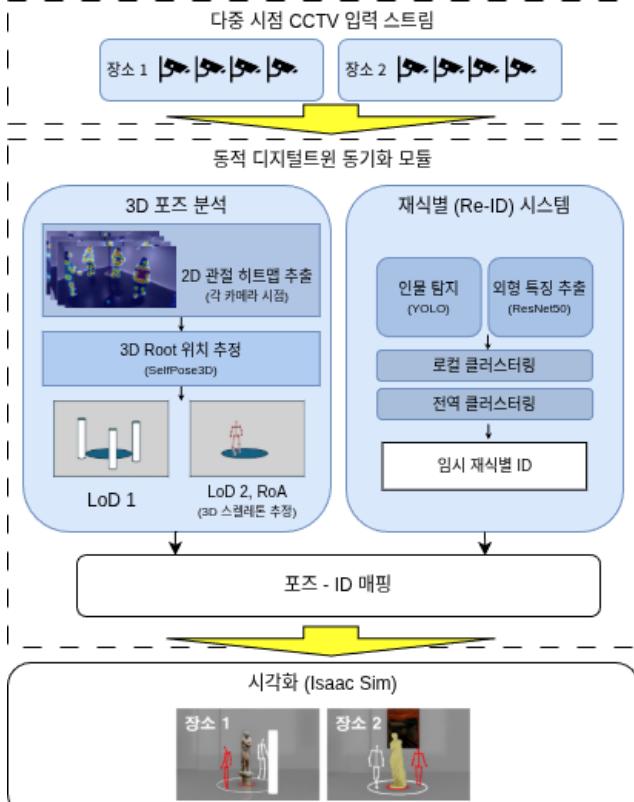


그림 1. 다중 시점 CCTV 영상을 입력으로 하는 동적 디지털 트윈 지능형 보안 인프라 시스템 작업 흐름도

그림 1은 PoC 전시 환경과 이에 대응하는 디지털 트윈 시각화 결과를 나타낸다. 시나리오 상에서 전시물 인근을 이동하는 일반 관람객들은 LoD 기준 영역의 외부 및 내부 위치에 따라 분석 수준이 차등적으로 적용되며, LoD 1 단계에서는 원통형 실린더, LoD 2 단계에서는 기본 스켈레톤 형태로 시각화 된다. LoD 경계는 흰색 원형으로 시각화 되며, 사용자 UI를 통해 실시간 조작이 가능했다. 시각화 된 인물 객체들은 별도의 장기 추적이 수행되지 않으며, PoC 환경 내에서 관람객의 이동 흐름과, 전시물 주변의 상호작용 등 환경 맥락을 표현하는데 활용되었다. 위험 후보로 설정된 관람객은 전시물에 접촉하기 이전까지는 다른 관람객들과 동일하게 실린더와 스켈레톤 형태로 표현되다가, 접촉 이후 LoD 3 단계로 전환되어 시각적으로 구분 가능한 경보 스켈레톤으로 표현되어 추적이 활성화된다. Re-ID 모듈을 통해 전역적 임시 추적 ID가 부여되며, PoC 시나리오 상에서는 해당 객체가 LoD 경계를 벗어나거나 기준 장소를 이탈하더라도 표현 수준과 추적 상태가 유지되고, 다른 장소에 재등장했을 때에도 동일 객체로 연속 인식되는 것을 확인하였다.



그림 2. PoC 전시환경 및 디지털 트윈 시각화 이미지

IV. 결론

본 연구는 다중 카메라 기반 3D 포즈 추정, 익명화 된 재식별, 그리고 디지털 트윈을 결합한 다중적 인식 구조를 통해, 공간 내 모든 사용자를 동일하게 감시하는 기존 보안 방식에서 벗어나 위험 행동에 한정된 선택적 인식 프레임워크를 제안하였다. 제안된 시스템은 지속적이고 반복적인 인식 작업을 자동화함으로써 인간 경비 인력의 인지적 부담을 경감할 수 있다. 제안된 구조는 향후 실시간 차리, 행동 패턴의 점진적 학습, 그리고 더 넓은 공간 규모로의 확장을 자연스럽게 수용할 수 있는 설계를 갖추고 있어, 지능형 보안 시스템이 보다 능동적이고 맥락 인지적인 방향으로 발전하기 위한 기반을 제공한다. 본 연구는 이러한 관점에서, 인공지능 기반 보안 인프라가 공공 공간 전반에 걸쳐 실질적인 안전성과 운영 효율을 동시에 향상시킬 수 있는 가능성을 제시한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원(IITP-2026-RS-2021-II211816)과 2026년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구(No. RS-2022-II220545, 지능형 디지털 트윈 연합 객체 구성 및 데이터 프로세싱 기술 개발)와 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받아 무인이동체원천기술개발사업을 통해 수행되었음.(RS-2020-NR117734)

참 고 문 헌

- [1] Donald, F.M. (2018). Information processing challenges and research directions in CCTV surveillance. *Cognition, Technology & Work*.
- [2] D. Kim, D. Lee and J. Kim, "Warning Zone Abnormal Behavior Detection: Pre-Alarms for Risk of Exhibit Damage," 2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom), Hong Kong, China, 2024, pp. 331–335, doi: 10.1109/MetaCom62920.2024.00062.
- [3] S Roka et al, A survey on abnormal behavior detection-based intelligence information video surveillance system using optimized machine learning methods
- [4] V. Srivastav, K. Chen, N. Paday, SelfPose3d: Self-supervised multi-person multiview 3d pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024, pp. 2502–2512.
- [5] M. Mittal et al., "Isaac Lab: A GPU-Accelerated Simulation Framework for Multi-Modal Robot Learning," *arXiv preprint arXiv:2511.04831*, 2025.