

제로샷 모델 기반 무인 매장 이상행동 탐지 융합 모델

이세훈*, 권기현**, 안준호***

Zero-shot Model-based Abnormal Behavior Detection Fusion Model in Unmanned Store

Sehun Lee*, Gihwon Kwon**, and Junho Ahn***

본 연구는 2025학년도 경기대학교 대학원 연구원장학생 장학금 지원에 의하여 수행되었음
이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신 기획평가원-학·석사연계 ICT핵심인재양성
지원을 받아 수행된 연구임(IITP-2025-RS-2024-00436954)

요약

본 논문에서는 무인 매장에서 발생하는 범죄 사고를 예방하기 위한 제로샷 모델 기반 이상행동 탐지 비전 융합 모델을 제안한다. 24시간 운영되는 무인 매장에서 파손, 폭행 등 범죄 상황이 발생할 수 있으며 빠른 대처가 유인 매장에 비해 느리다. 이를 해결하기 위해서, 사람들의 이상행동을 탐지하고 매장의 기물 및 상품 등을 파악하는 기술이 필요하다. 본 논문은 이상행동 탐지를 위해서 제로샷 기반 이미지 분할, 깊이 추정과 모션 캡처, 객체 탐지, 다중 객체 추적을 융합한다. 기물 및 상품 파악을 위해서 모든 객체를 학습한 탐지 모델이 필요하지만, 비용과 복잡도 면에서 어려운 점이 존재한다. 이를 극복하기 위해 제로샷 기반의 이미지 분할과 깊이 추정 알고리즘을 융합하여 미학습된 객체를 탐지할 수 있게 되었다. 제안한 모델로 이상행동 탐지를 실험한 결과로 정확도 0.919, 재현율 0.934, 정밀도 0.966, F1 점수 0.949의 이상행동 인식 성능을 보여주었다.

Abstract

This paper proposes a zero-shot - based vision model for abnormal-behavior detection designed to prevent criminal incidents in unmanned stores. Because such stores operate around the clock, vandalism or assaults may occur, and response times are typically slower than in manned stores. To address this issue, technology is required that can both detect customers' abnormal behaviors and monitor store fixtures and merchandise. The proposed approach fuses zero-shot based image segmentation, depth estimation and motion capture, object detection, and multi-object tracking to identify abnormal behavior. Although a fully supervised detector covering every possible object would be ideal for inventory monitoring, the associated cost and complexity are prohibitive. By combining zero-shot segmentation with depth estimation, our method can recognize previously unseen objects, thereby improving detection coverage for all store items. Experimental results show that the proposed model achieves an accuracy of 0.919, recall of 0.934, precision of 0.966, and an F1 score of 0.949 in abnormal-behavior recognition.

Keywords

zero-shot model, abnormal behavior detection, motion capture, object detection, multi-object tracking, unmanned store

* 경기대학교 컴퓨터공학과 석사과정

- ORCID: <https://orcid.org/0009-0006-0293-7544>

** 경기대학교 AI컴퓨터공학부 교수

- ORCID: <https://orcid.org/0000-0002-8221-4939>

*** 경기대학교 AI컴퓨터공학부 부교수(교신저자)

- ORCID: <https://orcid.org/0000-0003-2916-6410>

· Received: Apr. 29, 2025, Revised: May 19, 2025, Accepted: May 22, 2025

· Corresponding Author: Junho Ahn

Dept. of AI Computer Science and Engineering, Kyonggi University, 154-42

Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, South Korea

Tel.: +82-31-249-9731, Email: jha@kyonggi.ac.kr

I. 서론

코로나19 이후, 무인 매장의 수가 증가하고 있다. 2023년 소방청에 따르면, 사진관, 세탁소, 아이스크림 가게, 밀키트 매장, 스터디 카페, 룸카페 등 6종류의 무인 매장이 전국 6,323곳에 이르는 것으로 조사되었다[1]. 하나금융경영연구소의 조사에 따르면, 2023년 대표 편의점 4사의 무인 매장은 3,310개로 2020년 대비 6배 증가한 것으로 조사되었다[2]. 기존의 무인 매장 외에도 노래방, 꽃 가게, 문구점, 의류매장 등 다양한 무인 매장이 등장하는 추세이다. 신한카드 빅데이터연구소의 조사에 따르면, 2023년 기준 전년도 대비 무인 형태의 카페, 문구점, 아이스크림 매장, 세탁소 등에서의 이용 건수가 급증한 것으로 확인되었다[3]. 이러한 무인 매장의 수와 이용 건수 급증의 이유는 이용자들이 원하는 시간에 이용할 수 있고 인건비 절감으로 인해 저렴한 상품과 서비스를 제공한다는 장점으로 분석된다.

늘어나는 무인 매장 이용 건수 증가와 같이 무인 매장 관련 범죄도 증가하고 있다. 경찰청 통계에 따르면, 2022년 무인 매장 절도 건수는 2021년 대비 71% 증가하여 6,018건으로 조사되었다[4]. 무인 매장에서 발생하는 범죄의 종류는 절도, 분실·도난 카드의 부정 사용, 점유이탈물 횡령, 재물 손괴, 무전 취식뿐만 아니라 폭행이나 성폭행과 같은 강력 범죄도 발생하고 있다. 대기업 계열의 대표 무인 편의점 4사는 범죄에 대한 자체 대응 방안으로 대응 중이지만 자영업자의 무인 매장은 폐쇄회로(CCTV) 외에는 보안 방법이 취약하여 범죄의 표적이 될 수 있다. 또한 CCTV의 경우에는 무인 매장 점주가 영상을 수시로 확인해야 하는 불편함이 존재한다.

본 논문에서는 실시간 무인 매장 이상행동 탐지하기 위한 객체 탐지, 객체 추적, 모션 캡처, 제로샷 모델 기반 이미지 분할, 깊이 추정을 통해 비전 기반 융합 모델을 제안한다. 논문에서 정의하는 이상행동은 사용자의 반복되는 정상 패턴에서 비정상적으로 일어나는 행동 패턴이며, 급격한 동작 변화(달리기, 점프, 실신)와 일반적이지 않은 자세 그리고 물건 파손 및 투척 등을 의미한다. 제안한 모델은 다음과 같은 3가지의 장점이 있다. 첫 번째로는 제

로샷 기반 탐지 방식을 통해 추가 학습 비용을 최소화한다. 물건 파손 및 투척 등의 객체가 수반되는 이상행동을 인식하기 위해서 다양한 객체 탐지가 필요하다. 하지만 모든 객체를 학습한 모델을 구축하는 것은 상당한 비용과 복잡성을 요구하기에 실제 적용에는 한계가 존재한다. 이를 위해 지도학습 기반 객체 탐지로 주요 물체를 탐지하고, 제로샷 기반 이미지 분할, 깊이 추정을 통해 다양한 미학습 객체까지 인식한다. 두 번째로는 객체 추적을 활용해 혼잡 및 가려짐 환경에서도 객체 탐지 정확도를 유지한다. 혼잡한 상황에서 발생하는 객체 폐색으로 인한 탐지 ID 전환 문제를 해결하고 사용자의 이동량을 탐지할 수 있다. 세 번째로는 모션 캡처와 깊이 추정을 융합한 정밀 행동 분석이 가능하다. 모션 캡처를 이용하여 정상적인 행동 패턴에 대한 각속도를 수집한다. 이를 기반으로 사용자의 관절 각도를 인식하여 이상행동을 판단하는 로직을 개발한다. 또한 객체가 수반된 이상행동을 판단하는 로직을 분할 마스크 결과와 깊이 맵을 이용해 개발한다.

본 연구에서는 제로샷 기반 이상행동 탐지 비전 융합 모델을 개발하여 단순 이상행동과 객체가 수반된 이상행동을 구분하여 탐지할 수 있다. 이는 이상행동 인식의 정확도를 높일 수 있게 된다. 제안한 융합 모델은 무인 매장 업주가 매장을 효과적으로 관리할 수 있게 해주며, 범죄를 예방하고 빠르게 대처할 수 있도록 이상행동 인식할 수 있다.

본 논문의 2장에서는 객체 추적과 탐지, 모션 캡처, 그리고 제로샷 기반 이미지 분할, 깊이 추정을 활용한 연구에 관해 서술한다. 3장에서는 개별 알고리즘과 융합 모델을 설명하며, 4장에서는 개별 알고리즘과 융합 모델에 대한 실험 방법과 실험 결과를 서술한다. 마지막 5장에서는 본 연구의 결론 및 활용 방안, 향후 연구의 방향성을 서술한다.

II. 관련 연구

2.1 제로샷 모델

제로샷 모델 기반의 이미지 분할과 깊이 추정 기술은 산업 및 이상 탐지 분야에서 연구되고 있다.

Meta AI에서 개발한 제로샷 기반 이미지 분할 알고리즘을 활용해 가금류(닭)의 이상행동 패턴을 파악한 연구도 있다. 이 연구는 닭의 무게를 추정하고 행동 패턴을 분석하여, 제로샷 모델이 축산업에서도 활용될 수 있음을 입증한다[5]. 실제 산업 환경에서 수집한 공정 이미지를 통한 제로샷 기반 이상 탐지 연구도 존재한다[6]. 제로샷 모델의 인코더와 디코더를 경량화하여 이미지에서 정상과 이상 특징을 추출하고 이를 분류한다. 실시간성을 높여 드론 기반 철도 선로 위험 요소 인식에 제로샷 모델을 적용한 연구는 이미지 분할을 통해 객체를 탐지하고 철도 안전 검사를 수행한다[7]. 지도학습과 제로샷 모델을 융합한 Fusion Vision 방식을 통해 높은 3D 객체 탐지 성능을 구현한 연구도 있다[8]. 이 연구는 객체 탐지를 선행하고 제로샷 기반 이미지 분할을 진행하여 정밀한 객체의 경계 마스크를 생성한다. 제로샷 기반 깊이 추정 알고리즘도 이상 탐지 분야에서 주목받고 있다. 단안 카메라 기반 깊이 추정 모델(MiDaS)을 활용하여 다이아몬드 표면의 이상을 탐지한다. 이 연구는 RGB 영상으로 생성한 깊이 맵을 이용해 합성 다이아몬드의 품질을 판별한다[9]. 또한 RGB와 깊이 정보를 함께 사용하는 3D 표면 이상 탐지 기법인 3DSR(Dual Subspace Reprojection)을 개발한 연구는 기존 한계를 극복하며 MVTec3D 벤치마크에서 뛰어난 정확도와 속도를 기록한다[10]. 그러나 깊이 추정 기반 알고리즘은 표면 이상 탐지에 집중되어 있으며, 이상행동 탐지 분야에서는 상대적으로 활용이 적은 상황이다. 이에 본 논문은 제로샷 기반 깊이 추정 기술을 이상행동 탐지에 접목하여 이상을 정밀 탐지하고자 한다.

2.2 다중 객체 추적 및 탐지

범죄 증가로 인해 지능형 CCTV의 필요성이 부각되면서 CCTV를 활용한 이상 패턴 인식 및 객체 탐지 연구가 활발히 진행되고 있다. 객체 탐지와 모션 캡처를 활용한 클라우드 연계형 지능형 에지 CCTV 시스템이 개발되었으며, 이는 분산 에지 클라우드 기술을 기반으로 행동 패턴을 인식한다

[11]. 객체 탐지 모델과 CNN 기반 기술을 융합하여 무기 탐지 및 감정 분석을 실시간으로 수행하는 연구도 진행되었으며, 저해상도 CCTV 영상을 Real-ESRGAN 알고리즘을 통해 고해상도로 개선해 탐지 정확도를 높였다[12]. 실시간 이상행동을 탐지하기 위해 경량화된 맞춤형 CNN 아키텍처를 개발한 연구도 이루어졌으며, 이는 실시간성 확보와 이상행동 인식 모두를 가능하게 했다[13]. Inception-v3 기반 CNN 모델을 이용하여 비디오 및 이미지 데이터에서 폭행, 넘어짐 등과 같은 이상행동의 패턴을 학습한 연구도 있다[14]. 학습한 모델은 이미지 내에서도 정상행동과 이상행동 분류가 가능하도록 설계되었다. 주거 공간에서 맞춤형 행동 패턴을 분석하여 이상행동을 인식하는 시스템도 개발되었으며 이 시스템은 딥러닝 기반 행동 분류와 시퀀스 패턴 분석을 통해 실시간으로 이상행동을 탐지할 수 있도록 설계되었다[15]. 반려묘의 행동 패턴 분석 및 물리적 질병을 예측하기 위해 객체 추적을 활용한 연구도 있다[16]. 실시간 객체 탐지와 추적을 융합해 반려묘 간의 ID Switching 없이 추적할 수 있었다. KCF 추적기를 활용해 보행자의 이상행동을 실시간으로 추적하는 연구에서의 이 알고리즘은 템플릿 방식으로 빠른 처리 속도를 제공하지만 스케일 변화와 보행자 가려짐에 따른 드리프트 문제에 어려운 한계가 있다[17][18]. 이러한 문제를 극복하기 위해 본 논문에서는 ByteTrack을 도입하여 객체 간의 폐색 상황에서도 안정적인 추적과 ID 유지가 가능하게 하였다. 이외에도 Multi-Object Tracking(MOT) 기술을 활용하여 지능형 실내 위치 정보 분석 시스템을 제안한 연구도 진행되고 있다[19].

2.3 모션 캡처

본 논문에서는 무인 매장 사용자의 자세 추정을 위해 모션 캡처를 사용한다. 앞선 연구로는 모션 캡처를 활용해 데이터를 전처리 작업을 하고 슬라이딩 윈도우 방식을 이용해 GRU 학습 데이터를 생성 및 분석해 독거노인의 이상행동을 인식하는 연구가 있다[20]. 먼저 모션 캡처를 통해 발작, 실신, 무반응

등의 행동 패턴을 학습하고, 학습된 패턴이 인식되면 이상행동으로 판단하여 자동 신고할 수 있도록 했다. 다른 연구로는 무인 매장에서 객체 탐지만으로는 인식하기 어려운 사용자의 손 모양의 탐지를 추가한 연구도 있다[21]. 모션 캡처를 활용해 사람의 손 모양 패턴을 K-NN 알고리즘으로 학습시켜, 손을 중심으로 존재하는 이상행동을 인식하도록 설계되었다. 이외에도 운전자의 이상행동 탐지를 위한 객체 탐지, 모션 캡처 알고리즘 융합 연구도 있다 [22]. 운전 중 흡연, 음료 섭취 등을 탐지하기 위해 운전자와 물체의 상호작용을 탐지한다.

III. 제안한 융합 모델

3.1 이상행동 탐지

여러 접근 방식으로 이상행동을 인식하는 연구들이 진행되었다[11-17]. 앞선 연구들은 특정 모션 캡처에 중점을 두거나 특정 객체 탐지에 집중하여 사람의 행동 패턴을 파악해 이상행동을 인식했다. 그러나 이러한 연구들은 다양한 객체와 관련된 이상행동 특히 객체가 수반된 이상행동을 정확히 파악하는 데 한계가 있다. 모션 캡처 알고리즘이나 행동 패턴 분석으로 사람의 발차기를 인식할 수는 있지만, 그것이 단순한 발차기인지 물건을 부수는 발차기인지까지는 알 수가 없다. 또한 총이나 칼과 같은 특정 객체 탐지에만 집중하면 예상치 못한 둔기나

일상생활에서 우발적으로 집은 물체(예: 병, 우산, 의자 등)가 무기로 사용될 경우는 인식할 수가 없다. 이러한 한계를 어떠한 객체가 수반되어도 이상행동을 탐지하는 융합 모델을 제안한다.

먼저, 객체 탐지를 활용해 주요 객체(사람, 키오스크)를 인식한다. 사람이 인식되면 모션 캡처와 다중 객체 추적을 통해 자세 추정과 이동량 파악 및 추적을 통해 이상행동 여부를 판단하게 된다. 이때, 파운데이션 기반 이미지 분할과 깊이 추정을 활용하여 학습되지 않은 객체를 인식하고, 이를 기반으로 객체가 수반된 이상행동인지 아닌지를 인식하게 된다. 이 융합 모델은 주요 객체와 학습되지 않은 객체를 모두 인식할 수 있으며, 사람의 자세 추정과 이동량 파악 및 추적을 통해 정밀한 이상행동 인식이 가능하다. 이를 통해 각 알고리즘을 개별적으로 사용할 때 발생하는 이상행동 인식의 한계를 극복한 융합 모델을 제안한다.

그림 1은 본 논문에서 제안한 이상행동 인식 모델의 전체 구조를 나타낸다. 제로샷 기반 객체 탐지 모듈에서 미학습 객체를 탐지하고 깊이 추정을 통해 객체 간의 연관성을 알아낸다. 지도학습 기반 객체 탐지 모듈은 학습된 객체를 탐지하며, 모션 캡처와 객체 추적을 통해 이상행동 탐지를 진행한다. 이때 정상행동은 먼저 결과로 나오며, 모듈에서의 결과를 토대로 객체가 수반된 이상행동과 객체가 수반되지 않은 이상행동 그리고 정상행동을 분류한다.

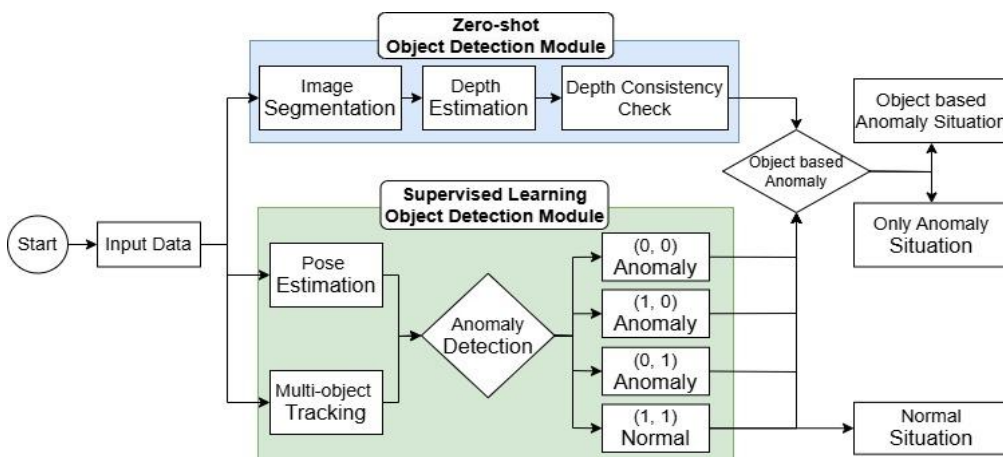


그림 1. 다중 비전 식별을 융합한 제로샷 모델 기반 이상행동 인식 모델 구조

Fig. 1. Zero-shot model-based abnormal behavior recognition model architecture fusing multi-vision identification

3.2 제로샷 모델 기반 객체 탐지

그림 2는 제로샷 모델을 통한 모듈의 과정을 나타낸 것이다. 먼저 입력된 이미지의 전체 이미지 분할[23]과 깊이 추정[24]을 진행한다. 이를 통해 학습되지 않은 객체도 마스크 데이터로 탐지가 되며, 각 객체 마스크의 깊이 데이터를 추출할 수 있다. 각 객체 마스크의 깊이 데이터를 중앙값으로 계산 후 사람 마스크의 깊이 데이터(식 (1))와 비교 연산한다(식 (2)). 유사한 중앙값을 가진 객체 마스크가 사람과 연관 있음을 알아낸다(식 (3)). 연관 있음이 밝혀진 객체는 객체가 수반된 이상행동 탐지 여부에 사용된다.

$$D_P = \{D(i,j) | (i,j) \in M_P, M_P(i,j) = 1\} \quad (1)$$

$$median_P = median(D_P)$$

$$D_{sam}^{(k)} = \{D(i,j) | (i,j) \in M_{sam}^{(k)}, M_{sam}^{(k)}(i,j) = 1\} \quad (2)$$

$$median_{sam}^{(k)} = median(D_{sam}^{(k)})$$

$$|median_{sam}^{(k)} - median_P| \leq T_d \quad (3)$$

$$M_{filtered}^{(k)} = \{M_{sam}^{(k)} | |median_{sam}^{(k)} - median_P| \leq T_d\}$$

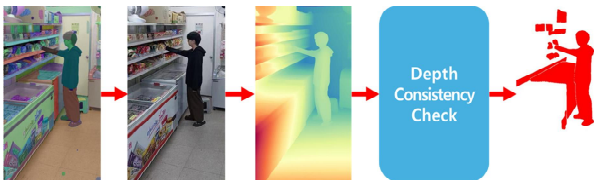


그림 2. 제로샷 기반 사람과 주위 객체 연관성 파악
Fig. 2. Zero-shot model-based human-surrounding object correlation analysis

사람이 객체가 수반된 이상행동을 저지른다고 하여도 물건을 던지지 않는 한 주위의 객체를 대상으로 하거나 근접한 객체를 사용한다. 이때 깊이 값 연산으로 사람과 유사한 객체만을 필터링함으로써 상호연관성이 있는 객체를 추출하여 객체가 수반된 이상행동 탐지 정확도를 높인다. 그림 2와 같은 제로샷 기반 객체 탐지 모듈 과정을 이용하면 이상행동 시 객체가 수반된 행동 여부를 파악할 수 있다.

3.3 지도학습 기반 객체 탐지

그림 3은 지도학습 기반 객체 탐지를 활용한 이상행동 탐지 모듈을 과정을 나타낸다. 먼저 학습된 모션 캡처[25]로 사람의 행동 패턴을 분석 및 탐지한다. 다중 객체 추적[26]을 통해 사람의 이동량을 분석 및 탐지한다. 각 알고리즘으로 이상 패턴을 분석하여 두 알고리즘의 결과를 융합하여 이상행동을 분류한다. 이상 패턴 탐지 시를 1, 정상 패턴 탐지 시 0으로 가정하여 각각 결과는 (1, 1), (1, 0), (0, 1), (0, 0)으로 산출한다.

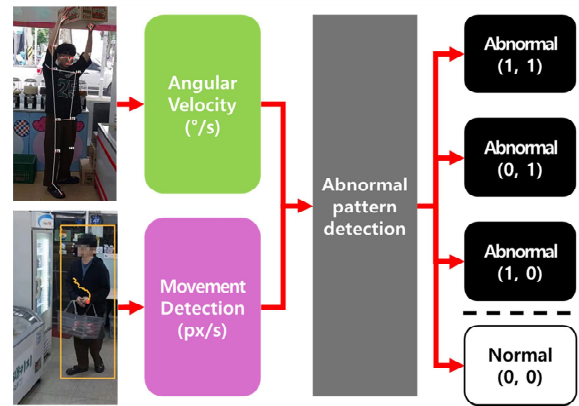


그림 3. 각속도 및 변화량 기반 이상 탐지
Fig. 3. Angular velocity and change-based anomaly detection

3.4 각속도 기반 이상행동 탐지

표 1은 이상행동 패턴 분석을 위해 관측한 주요 관절과 각속도 수치다. 발차기, 주먹질, 점프 등의 이상행동 데이터를 바탕으로 관절 각도 분석했다. 팔꿈치-어깨-골반으로 이어지는 겨드랑이 각도의 임계값은 120도, 어깨-골반-무릎으로 이어지는 골반 각도의 임계값은 73도, 골반-무릎-발목으로 이어지는 무릎 각도의 임계값은 77도로 패턴 분석했다.

표 1. 각속도 기반 관측된 이상 패턴

Table 1. Abnormal patterns detected from angular velocity data

Angle type	Abnormal value
Elbow-Shoulder-Pelvis	120°/s
Shoulder-Pelvis-Knee	73°/s
Pelvis-knee-ankle	77°/s

분석한 임계값을 토대로 모션 캡처를 활용해 무인 매장 내 사람의 이상행동을 탐지한다.

3.5 이동량 기반 이상행동 탐지

실내 무인 매장에서의 이상행동이 포함된 영상 데이터를 통해 정상행동과 이상행동의 차이를 분석한다. 그림 4는 정상행동 시 측정된 객체 추적 값(ID:4)과 이상행동(달리기, 점프, 실신 등) 시 측정된 객체 추적 값(ID:2) 중 달리기 시 측정된 수치다. 그림 4에서 y축은(px/s) 초당 픽셀 변화량을 나타내며 약 2300px/s로 표시된 ID:4의 값은 이상행동 시의 수치이며 100px/s 이하인 ID:2의 값은 정상행동 시 측정된 수치다. 이상행동과 정상행동의 객체 추적 변화량의 차이를 비교 분석한다. 이러한 분석을 통해 이상행동 시 객체 추적의 변화량을 940px/s로 설정하여 객체 추적을 통해 측정된 무인 매장 내 사람의 변화량을 바탕으로 이상행동 여부를 판별한다.

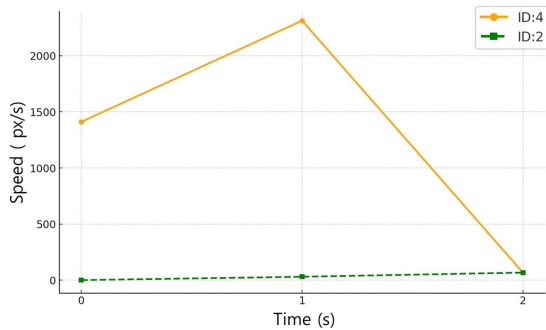


그림 4. 이상행동 추적량 분석
Fig. 4. Abnormal behavior tracking analysis

3.6 객체 수반 이상행동 탐지

제로샷 기반 객체 탐지 모듈과 지도학습 기반 객체 탐지 모듈의 출력값을 토대로 최종 이상 탐지를 진행한다. 지도학습 기반 객체 탐지 모듈에서의 출력값이 (0, 0)이면 정상 패턴으로 분류하며, 그 외의 출력값인 (0, 1), (1, 0), (1, 1)은 이상 패턴으로 분류한다. 그 후 제로샷 기반 객체 탐지 모듈에서의 출력과 비교 분석한다. 이상 패턴으로 분류된 시점(t)에서 ± 1 초(t-1, t, t+1)동안 사람과 연관 있는 객체 마스크가 존재함을 확인한다. 사람과 연관 있는 객체 마스크가 t-1~t+1 사이에 존재하였다면 객체가

수반된 이상행동(물품 훼손, 물품 던짐, 흥기 휘두르기 등)으로 탐지한다. 제로샷 기반 객체 탐지이므로 객체가 수반된 이상행동 간의 분류가 불가능하여도 추가적인 학습 없이 최종적으로 객체 자체가 수반된 이상행동임을 확인할 수가 있다.

IV. 모델 실험 및 분석

4.1 데이터 수집 및 실험 환경

실험 데이터는 약 5,000장의 사진과 100개의 영상을 실제 무인 매장 3곳에서 촬영했으며, 추가로 인터넷으로 수집했다. 지도학습 기반 탐지하는 주요 객체의 경우 직접 라벨링 작업 후 데이터 증강(회전 -7° , $+7^\circ$, 채도 조절 -18% , $+18\%$, 노이즈 추가 0.42%)을 진행하였다. 학습, 테스트, 검증 데이터셋 분리 비율은 주요 객체 데이터 내에서만 6:2:2 비율로 나누어 실험을 진행했다. 성능 비교를 위해 수집한 모든 사진과 영상은 640p, 720p, 960p 해상도로 전처리 과정을 진행했다. 실험은 NVIDIA® GeForce RTX 4090 GPU와 13th Gen Intel(R) Core(TM) i9-13900K 3.00 GHz CPU 하드웨어 환경에서 진행했으며, 소프트웨어 환경은 Python 3.8.18, Pytorch 2.3.1, CUDA 12.5를 사용하여 모델 학습과 실험을 수행했다. 모델의 탐지 성능을 평가하기 위해 Precision, Recall, Accuracy, F1 Score를 사용했다.

4.2 사용자 인접 객체 탐지 실험 결과

제안한 모델의 주변 객체 탐지를 위해 제로샷 기반의 SAM[23], Depth Anything[24] 알고리즘을 융합하였다. 제안한 제로샷 기반 융합 모델 모듈로 탐지된 주변 객체와 실제 사람이 상호작용이 가능한 객체를 비교해서 실험 결과를 표 2에 나타낸다. 학습되지 않은 객체에 관한 탐지 결과 모든 지표에서 0.8 정도의 근사치 값을 확인할 수 있었다.

표 2. 주변 객체 탐지 실험 결과
Table 2. Surrounding object detection experiment result

Precision	Recall	Accuracy	F1 Score
0.896	0.801	0.793	0.846

4.3 이상행동 탐지 실험 결과

제안한 모델의 이상행동 탐지를 위한 모션 캡처는 MediaPipe[25]를 사용했으며, 다중 객체 추적으로는 ByteTrack[26]을 사용했다. 표 3은 모션 캡처, 객체 추적 각각의 이상행동 탐지 성능과 두 알고리즘을 융합했을 때의 성능을 나타낸다. 융합 모델의 성능이 개별 알고리즘의 성능보다 높음을 확인하였다.

표 3. 이상행동 탐지 실험 결과
Table 3. Anomaly detection experiment results

Parameters Method	Precision	Recall	Accuracy	F1 Score
Motion capture	0.895	0.946	0.867	0.920
Object tracking	0.929	0.939	0.876	0.934
Fusion	0.943	0.951	0.968	0.948

4.4 상황 및 해상도 별 융합 모델 실험 결과

제안한 융합 모델의 성능 평가를 위해 수집한 데이터로 실험했다. 수집한 데이터는 객체가 수반된 이상행동, 수반되지 않은 이상행동, 정상행동 등 데이터를 수집하였다. 이를 통해 다중 비전 식별을 융합한 제로샷 모델 기반 이상행동 인식 실험하였다. 표 4는 수집한 데이터를 분류하여 제안한 모델의 성능을 나타낸다.

표 4. 상황별 융합 모델 실험 결과
Table 4. Fusion model experiment results by situation

Parameters Situation	Precision	Recall	Accuracy	F1 Score
walking	0.955	0.938	0.940	0.947
standing	0.951	0.939	0.962	0.944
sit down	0.927	0.931	0.943	0.929
lying	0.901	0.893	0.896	0.898
running	0.918	0.936	0.884	0.927
jump	0.912	0.920	0.908	0.916
kick	0.932	0.941	0.963	0.937
punching	0.954	0.934	0.955	0.943
swoon	0.887	0.907	0.911	0.898
object on jump	0.909	0.911	0.899	0.910
object kick	0.911	0.928	0.928	0.919
object punching	0.944	0.948	0.931	0.946
object throw	0.901	0.914	0.894	0.907

수집한 데이터 해상도 별 모델 성능은 표 5와 같다. 제안한 모델은 결과와 해상도 간의 상관관계 없이 일관된 성능을 나타낸다.

표 5. 해상도 별 융합 모델 실험 결과
Table 5. Fusion model experiment results by resolution

Parameters Resolution	Precision	Recall	Accuracy	F1 Score
640p	0.966	0.934	0.919	0.949
720p	0.959	0.940	0.917	0.949
960p	0.968	0.932	0.922	0.950

V. 결론 및 향후 과제

무인 매장에서의 안전과 보안을 위해 다양한 비전 알고리즘을 통한 이상행동 인식 융합 모델을 제안한다. 현재 무인 매장이 많아지면서 많은 범죄에 노출되고 있다. 무인 매장 내 절도, 폭행, 중범죄뿐만 아니라 실신 등의 안전사고에 대한 빠른 대처를 위한 이상행동 인식 융합 모델을 개발했다. 융합 모델에 대한 성능 결과는 정밀도, 민감도, 정확도, F1 score로 구했으며, 각각 0.966, 0.934, 0.919, 0.949의 결과가 나왔다. 또한 실제 무인 매장 영상을 수집하여 상황별 모델 비교 실험을 진행했다. 그 결과 13개의 상황에서 대부분 0.9 이상의 결과를 보였다.

제안한 융합 모델의 사용은 무인 매장에서 이상행동을 인식하고 이를 매장 주인에게 빠른 정보 전달이 가능할 것이다. 이는 무인 매장 운영에 비용적 부담과 사고 예방에 도움이 된다. 또한 다양한 무인 매장에서 사용이 가능한 범용성 있는 융합 모델로 활용성이 높다. 추후 본 연구는 다양한 무인 매장 종류에 맞게 맞춤형 이상행동 징후 패턴을 학습시켜 맞춤형 융합 모델로 개발할 예정이다. 또한 무인 매장에서 나아가 실내 CCTV가 있는 모든 환경 내에서의 이상행동 인식을 위한 모델 구축 및 파라미터 튜닝해 범용성을 높일 예정이다.

References

[1] National Fire Agency, Promotion of Fire Risk Assessment for Multi-Use Businesses, 2023. <https://www.nfa.go.kr/nfa/>. [accessed: Mar. 02, 2025]

- [2] Hana Financial Management Research Institute, Retail unmanned, critical point is coming, 2023. <https://www.hanaif.re.kr/main.do>. [accessed: Mar. 02, 2025]
- [3] ShinganCard BIG DATA, Trends in the number of unmanned store uses, 2024. <https://www.shinhancard.com/pconts/bdinfo/html/allview/bignsight.html>. [accessed: Mar. 02, 2025]
- [4] Korean National Police Agency, National Police Agency Crime Statistics, 2022. <https://www.police.go.kr/index.do>. [accessed: Mar. 02, 2025]
- [5] X. Yang, et al., "An innovative segment anything model for precision poultry monitoring", *Computers and Electronics in Agriculture*, Vol. 222, pp. 109045, Aug, 2024. <https://doi.org/10.1016/j.compag.2024.109045>.
- [6] C. Li, L. QI, and X Geng, "A SAM-guided Two-stream Lightweight Model for Anomaly Detection.", *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 21, No. 2, pp. 1-23, Jan. 2025. <https://doi.org/10.1145/3706574>.
- [7] S. LI, C. Yu, L. Chang, and X. Zhao, "Railway surrounding environment hazard detection based on fast SAM", *International Conference on Electrical and Information Technologies for Rail Transportation*. Singapore: Springer Nature Singapore, Beijing, China, pp. 644-656, Oct. 2023. https://doi.org/10.1007/978-981-99-9315-4_63.
- [8] S. E. Ghazouali, Y. Mhirit, A. Oukhrif, U. Michelucci, and H. Noura, "FusionVision: A comprehensive approach of 3D object reconstruction and segmentation from RGB-D cameras using YOLO and fast segment anything", *Sensors*, Vol. 24, No. 9, Apr. 2024. <https://doi.org/10.3390/s24092889>.
- [9] K. Fujiwara, "Anomaly detection for synthetic diamond grain using monocular depth estimation", *Computers & Industrial Engineering*, 2024, Vol. 188, No. 8, pp. 109904, Feb. 2024. <https://doi.org/10.1016/j.cie.2024.109904>.
- [10] V. Zavtanik, M. Kristan, and D. Skočaj, "Cheating Depth: Enhancing 3D Surface Anomaly Detection via Depth Simulation", *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision(WACV)*, Waikoloa, HI, USA, pp. 2164-2172, Jan. 2024. <https://doi.org/10.1109/WACV57701.2024.00216>.
- [11] Y. Kim, S. Jeong, S. Park, and Y. Park, "Threat Situation Determination System Through AWS-Based Behavior and Object Recognition", *KIPS Transactions on Software and Data Engineering*, Vol. 12, No. 4, pp. 189-198, Apr. 2023. <https://doi.org/10.3745/KTSDE.2023.12.4.189>.
- [12] S. Lee, S. Kim, C. Kang, J. Baek, and S. Jin, "Deep Learning Based Weapon Detection and Emotion Analysis Framework to Strengthen Public Safety", *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 28, No. 5, pp. 616-626, May 2024. <http://doi.org/10.6109/jkiice.2024.28.5.616>.
- [13] P. Y. Ingle and Y. Kim, "Real-time abnormal object detection for video surveillance in smart cities", *Sensors*, Vol. 22, No. 10, pp. 3862-3883, May 2022. <https://doi.org/10.3390/s22103862>.
- [14] S. A. Jebur, K. A. Hussein, and H. K. Hoomod, "Abnormal Behavior Detection in Video Surveillance Using Inception-v3 Transfer Learning Approaches", *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, Vol. 23, No. 2, pp. 210-221, Jun. 2023. <https://doi.org/10.33103/uot.ijccce.23.2.16>.
- [15] S. Choi, C. Kim, Y. S. Kang, and S. Youm, "Human behavioral pattern analysis-based anomaly detection system in residential space", *The Journal of Supercomputing*, Vol. 77, pp. 9248-9265, Feb. 2021. <https://doi.org/10.1007/s11227-021-03641-7>.
- [16] S. Han and D. W. Park, "Cat Behavior Pattern Analysis and Disease Prediction System of Home

- CCTV Images using AI", Journal of the Korea Institute of Information and Communication Engineering, Vol. 26, No. 9, pp. 1266-1271, Sep. 2022. <https://doi.org/10.6109/jkiice.2022.26.9.1266>.
- [17] D. Kim and S. Park, "Real-time Abnormal Behavior Analysis System Based on Pedestrian Detection and Tracking", Korea Institute of Information and Communication Engineering Spring Conference, Yeosu, Korea, Vol. 25, No. 1, pp. 25-27, May 2021. <https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE10565895>.
- [18] S. Sun and E. Lee, "Aerial Target Tracking in a Cloud Clutter Environment using Multi-sensor Information", The Journal of Korean Institute of Information Technology, Vol. 19, No. 4, pp. 75-82, Apr. 2021. <https://doi.org/10.14801/jkiit.2021.19.4.75>.
- [19] J. Jang, M. Seon, S. Nam, J. Choi, and K. Cha, "Design of An Intelligent Indoor Location Analysis System Using Multi-Object Tracking", Journal of Digital Contents Society, Vol. 23, No. 8, pp. 1519-1526, Aug. 2022. <https://doi.org/10.9728/dcs.2022.23.8.1519>.
- [20] C. Kwak, Y. Kim, J. Lee, J. Jung, and C. Yoo, "Deep Learning-Based Abnormal Behavior Detection and Monitoring Service for the Senior Living Alone", Proceedings of KIIT Conference, Jeju, Korea, pp. 461-465. Jun. 2022.
- [21] Y. Choi, J. Na, and J. H. Ahn, "Intelligent Motion Pattern Recognition Algorithm for Abnormal Behavior Detections in Unmanned Stores", Journal of Internet Computing and Services(JICS), Vol. 24, No. 6, pp. 73-80, Dec. 2023. <https://doi.org/10.7472/jksii.2023.24.6.73>.
- [22] X. Cai, S. Zhou, J. Yao, P. Cheng, and Y. Hu, "A Driver Abnormal Behavior Detection Method Based on Improved YOLOv7 and OpenPose", Advanced Intelligent Computing Technology and Applications. (ICIC), Vol. 14090, pp. 239-250, Aug. 2023. https://doi.org/10.1007/978-981-99-4761-4_21.
- [23] A. Kirillov, et al., "Segment anything", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, pp. 4015-4026. Oct. 2023. <https://doi.org/10.1109/ICCV51070.2023.00371>.
- [24] L. Yang, et al., Depth anything: Unleashing the power of large-scale unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 10371-10381, Sep. 2024. <https://doi.org/10.1109/CVPR52733.2024.00987>.
- [25] C. Lugaresi, et al., "Mediapipe: A framework for building perception pipelines", arXiv preprint arXiv:1906.08172, Jun. 2019. <https://doi.org/10.48550/arXiv.1906.08172>.
- [26] Y. Zhang, et al., "Bytetrack: Multi-object tracking by associating every detection box", Computer Vision – ECCV 2022, Tel Aviv, Israel, pp. 1-21, Oct. 2022. https://doi.org/10.1007/978-3-031-20047-2_1.

저자소개

이 세 훈 (Sehun Lee)



2025년 2월 : 경기대학교
인공지능전공(학사)
2025년 3월 ~ 현재 : 경기대학교
컴퓨터과학과 석사과정
관심분야 : 컴퓨터 비전, 딥러닝,
인공지능

권 기 현 (Gihwon Kwon)



1985년 2월 : 경기대학교
전자계산학과(이학사)
1987년 8월 : 중앙대학교
전자계산학과(이학석사)
1991년 2월 : 중앙대학교
전자계산학과(공학박사)
1999년 9월 ~ 2000년 8월 : 미국

Carnegie Mellon University 전산학과(방문 교수)
2006년 1월 ~ 2007년 2월 미국 Carnegie Mellon
University 전산학과(방문 교수)
2014년 4월 ~ 2016년 2월 : 한국정보과학회
소프트웨어공학소사이어티 회장
1991년 2월 ~ 현재 : 경기대학교 AI컴퓨터공학부 교수
2021년 4월 ~ 현재 : 경기대학교 SW중심대학 총괄책임자

안 준 호 (Junho Ahn)



2006년 2월 : 홍익대학교
컴퓨터공학과(공학사)
2008년 2월 : 연세대학교
컴퓨터공학과(공학석사)
2013년 12월 : University of
Colorado Boulder
컴퓨터과학과(공학박사)

2013년 12월 ~ 2017년 1월 : 한국전자통신연구원(ETRI)
선임연구원
2017년 2월 ~ 2023년 2월 : 한국교통대학교
컴퓨터공학전공 부교수
2023년 3월 ~ 현재 : 경기대학교
AI컴퓨터공학부 부교수