다중 CCTV 내 객체 트래킹을 통한 고객 동선과 체류 구역 분석 시스템

김창현, 김종각, 정인준, 이재웅, 김동균*

경북대학교

kch0536@knu.ac.kr, zx58000@knu.ac.kr, jijun897@knu.ac.kr, thska12@knu.ac.kr, dongkyun@knu.ac.kr*

Analysis system that tracks customer movements and locations in multiple CCTV

Kim Chang Hyeon, Kim Jong Gak, Jeong In Jun, Lee Jae Woong, sfsdffsfdfsdsdf, Kim Dong Kyun*

Kyungpook National Univ.

요 약

본 논문은 다중 CCTV 데이터 내 객체 트래킹을 통해 오프라인 매장 운영에 기여할 수 있는 시스템을 제안한다. 이 시스템은 YOLO 기반 객체 탐지, 실시간 다중 객체추적을 위한 DeepSORT, 다중 CCTV 영상 내 식별된 객체 간 일관성을 유지하기 위해 사전학습된 OSNet을 통합하여 효과적으로 객체를 탐지하고 고객의 동선과 구역 내 체류 시간을 추적한다. 이러한 시각화 및 수치 데이터를 바탕으로 고객의 상품의 선호도와 매대 배치에 따른 영향성을 분석할 수 있다. 이 연구는 비디오 객체 분석 기술의 활용 방향성과 기술의 발전에 기여하며, 다양한 소매업에서의 기존 설비를 유지 또는 적은 추가 투자비용으로 실질적인 경제적 가치를 제공할 수 있음을 보여준다.

Ⅰ. 서 론

2020년대에 접어들며, 코로나19 팬데믹과 인공지능(AI) 기술의 급격한 발전으로 인해 온라인 시장이 큰 성장을 이루게 되었다. 그럼에도 오프라인 매장만이 제공하는 소비자의 경험과 직원이 제공하는 서비스 등으로 인해 매장에서 적용 가능한 다양한 AI 활용 연구가 지속적으로 이루어지고 있다. 그중 CCTV를 활용한 비디오 분석 기술이 크게 대두되고 있다. 다양한 종류의 비디오 분석 기술들은 자동화된 모니터링 및 분석, 실시간 상황 판단, 효율성 증대 등의 혁신적인 역할을 가능하게 한다.

기존의 CCTV 데이터 분석 연구들은 주로 보안 목적과 이와 연관된 특정 행위 감지에만 초점을 맞춘 개별 모델로서, 오프라인 비즈니스 운영 및 마케팅 전략 수립에 직접적으로 활용하기 어렵다[1]. 더불어, 매장의 형태에 따라 단일 CCTV 데이터의 개별적 분석만으로는 유기적인 객체의 흐름을 판단하기 힘들다는 명확한 단점을 지닌다. 또한, 매장이 혼잡한 경우에도 객체의 탐지가 효율적으로 이루어지는가가 성능에 있어 무엇보다 중요하다. 본 연구에서는 이러한 기존 한계를 극복하기 위한 시스템을 제안한다. 제안된 모델은 비디오 분석 성능을 향상시키고 매장 운영에 도움이 될 만한 실질적 데이터를 제공하기 위해 다음과 같은 핵심 기술을 통합한다: 대규모 실시간 영상 내 고객의 이동이 많은 상황에서도 높은 처리 수준을 위해 YOLOv12 기반 객체 탐지를 활용하여 실시간성과 안정성을 확보한다. 또한, DeepSORT 알고리즘으로 객체별 ID 정보를 유지하고 동선을 추적하게 했으며 이들 영상 간의 동기화된 객체 추적을 위해 옴니스케일 네트워크(OSNet) 특징 추출을 추가했다. 이러한 시스템을 통해 유동량이 많은 매장 내에서 각 고객을 사각지대 없이 꾸준하게 추적함으로써 고객의 동선 및 구역 내 체류 시간을 토대로 상품의 선호도와 매출 영향을 분석할 수 있는 토대를 마련한다.

Ⅱ. 관 련 기 술

1. YOLO

YOLO(You Only Look Once) 알고리즘은 실시간 객체 탐지(Object Detection) 분야에서 널리 사용되고 있는 딥러닝 알고리즘이다. 기존 R-CNN 계열의 복잡한 파이프라인을 단순화하여 단일 신경망을 통해 객체의 위치(Bounding Box)와 클래스를 동시에 예측하는 새로운 접근법을 제안한다. 이러한 예측 과정에서 단일 컨볼루션 신경망이 이미지 전체를 End-to-End로 처리하며 분류, 위치 추정, 신뢰도 추정을 동시에 수행한다. 이는 속도와 정확성의 균형이 뛰어나서 실시간 시스템에 적합하며, 이미지 전체의 문맥 정보를 활용하여 배경 오류(false positive)를 줄이고 일반화 성능이 뛰어나다는 특징이 있다[2].

2. DeepSORT

DeepSORT(Deep Simple Online and Realtime Tracking)는 SORT의 속도와 단순성을 유지하면서, 실시간 다중 객체 추적(Multi-Object Tracking, MOT) 문제에서 프레임 간 객체의 ID를 일관되게 유지하며 추적할 수 있는 알고리즘이다. 외부 객체 탐지의 결과물에서 사전 학습된 CNN기반 임베딩 네트워크를 통해 appearance feature vector를 추출한다. Kalman Filter를 기반으로 추적된 객체의 위치 상태를 예측하며, 두 가지의 거리 지표(운동 기반의 Mahalanobis Distance와 CNN 특징 벡터 간의 유사도 기반의 Cosine Distnace)를 혼합하여 거리 행렬을 계산한다. 이후 헝가리안 알고리즘(Hungarian Algorithm)의 최소 비용 매칭을 통해 탐지 결과와 기존 트랙 간의 최적 쌍을 결정한다[3].

3. OSNet

단일 영상과 달리, 각기 다른 구도에서 촬영된 영상 내의 동일 객체를 찾아내기 위한 문제서 DeepSORT만의 적용은 appearance embedding이 일관성을 잃게 되어 한계로 작용한다. 사람 재식별(Person Re-ID) 문제에서 다양한 공간 스케일의 시각 특징(Omni-scale features)을 효과적으로 학습하는 새로운 CNN 아키텍처인 OSNet(Omni-Scale Network)은 높은 정확도를 제공하여 이러한 문제 상황을 만족한다. OSNet에서 다중 스케일 스트림으로 구성된 Omni-Scale Residual Block을 통해 각 스트림은 서로 다른 receptive field에서 특징을 감지한

다. 이후 통합된 AG(Aggregation Gate)를 통해 입력에 따라 채널별 가중치를 부여하여 다중 스케일 특징을 동적으로 융합함으로써 각 입력에 가장 적합한 스케일을 적응적으로 강조할 수 있다. OSNet은 pointwise convolution과 epthwise convolution의 조합을 통해 모델의 복잡도를 줄이고 과적합을 방지하는 효율적인 처리 구조를 지닌다[4].

Ⅲ. 실 험

1. 제안 시스템

제안된 모델은 비디오 분석 성능을 향상시키고 매장 운영에 도움이 될 만한 실질적 데이터를 제공하기 위해 YOLOv12 모델과 DeepSORT 및 OSNet 알고리즘을 결합한 워크플로우를 가진다. 사전에 영상 내 매대, 카운터, 출입구 등의 영역을 지정한다. 첫 번째로 YOLOv12 모델을 사용하여 추출된 프레임에서 사람의 위치와 크기를 감지하며, 두 번째로 DeepSORT 알고리즘으로 감지된 사람들을 프레임 간에 추적하여 동일한 사람에게 고유한 ID를 부여한다. DeepSORT에 의해 추적된 각 사람의 바운딩 박스 영역에서 사전 학습된 OSNet을 사용하여 사람의 외형적 특징 벡터(feature vector)를 추출하고 다양한 스케일의 특징을 효과적으로 학습함으로써 사람 재식별의 정확도를 높이는 데 기여한다.

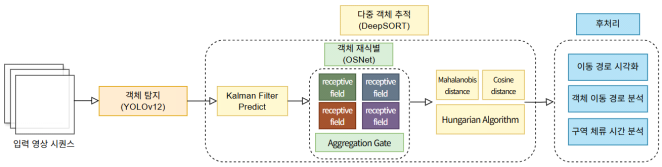

그림 1 . 시스템 흐름도

그림2, 3 과 같이 후처리 과정에서 각 객체의 이동경로와 미리 정해둔 구역별 체류 시간을 분석하여 기록하고 시각화한다.

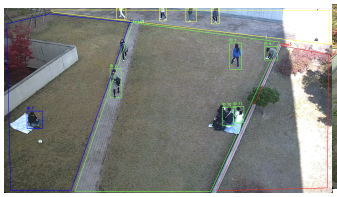

그림 2

그림 3

2. Implementation Details

이 모델의 학습 과정에서 이미지 크기는 1024로 조정하여 작은 객체의 탐지에도 유리하게 설정했으며, 실험에서 학습 속도 2e-4와 가중치 감소 0.1의 AdamW 최적화를 사용했다. AutoAugment를 사용하여 데이터의 자동증강과 0.5의 Mosaic, 0.4의 Random Erasing 및 Copy-Paste를 통해 데이터를 증강시키고 HSV 및 fliplr과 scale을 조정하는 기하학적 변형을 가하여 다양한 상황을 가정했다. 우리는 100의 에포크와 16의 배치 크기를 갖는 Linux 5.4.0-144-generic 및 PyTorch 2.5.1+cu121 개발환경에서 NVIDIA RTX A6000 GPU를 사용하여 모델을 훈련시켰다.

3. Datasets

Keio DBA Team이 Roboflow Universe에 공개한 CrowdHuman 데이터셋은 복잡한 군중 환경에서의 사람 탐지 성능을 향상시키기 위해 구성된 객체 탐지용 데이터셋이다. 총 9,285장의 다양한 밀집도와 가림 현상이 포함된 군중 이미지이며 각 사람 인스턴스에 대해 전체 신체, 가시 영역, 머리 등 다양한 바운딩 박스 주석이 제공되어, 다양한 수준의 사람 탐지 및 분석이 가능하다[5].

4. CrowdHuman 데이터셋에서 모델 평가

CrowdHuman 데이터셋은 다양한 실제 환경에서의 사람 탐지 성능을 평가하고 향상시키는 데 중점을 두고 있다. 그림4 와 같이 모델은 훈련과 검증 데이터에서 정밀도 80%, 평균 정밀도 mAP50에서 약 73%, 평균 정밀도 mAP50-95에서 약 47%라는 수치를 보여주어 높은 성능을 제공한다. Recall 값은 63%로 작거나 가려진 객체에 대해 상대적으로 낮은 성능을 보여준다.

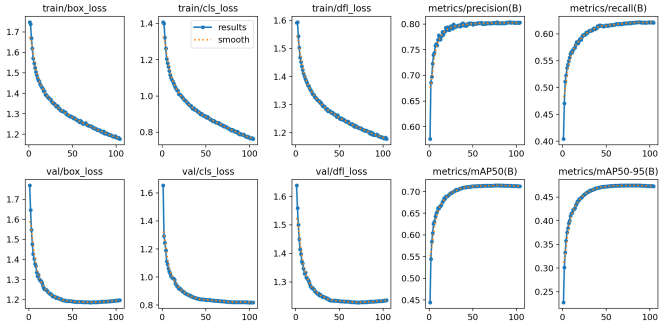

그림 4. CrowdHuman 데이터셋에서의 모델 성능

Ⅳ. 결 론

본 논문의 시스템은 매장 내 다중 CCTV 환경에서 비디오 데이터 내의 객체 특징을 구분하고 추적하여 고객의 동선을 파악하고 구역 내 체류 시간을 바탕으로 상품의 선호도와 매출 영향을 분석할 수 있다. 그러나 실제 매장 환경에 가까운 데이터셋을 사용하지 않아, 각 환경에 적용 시에 발생할 수 있는 성능 저하 등의 대두될 문제점을 다루지 못한다는 한계가 있다. 본 연구는 비디오 객체 분석 기술의 활용 방향성과 발전에 기여하며, 다양한 소매업에서의 기존 설비 또는 적은 추가 투자비용으로 실질적인 경제적 가치를 제공할 수 있음을 보여준다. 향후 제안된 시스템 내에 모션 인식이나 물품 데이터베이스와의 통합 등으로 확장하여 실제 환경에서 성능을 유지면서도 많은 기능을 제공하는 연구를 지속할 필요성이 있다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학 사업의 연구결과로 수행되었음"(2021-0-01082)

참 고 문 헌

[1] 곽내정, 이병엽. "딥러닝을 활용한 비디오 감시 시스템의 비정상적 인간 행동 탐지 연구 - 개요" International Journal of Contents 20, no.4 (2024): 84-95.doi: https://doi.org/10.5392/IJoC.2024.20.4.084

[2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric." 2017 IEEE international conference on image processing (ICIP). IEEE, 2017.

[4] Zhou, Kaiyang, et al. "Omni-scale feature learning for person re-identification." Proceedings of the IEEE/CVF international conference on computer vision. 2019.

[5] Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.