

객체 탐지를 활용한 스마트 모니터링 시스템

강민교*, 남현우, 강정완, 조용진, 김동재, 이재욱

부경대학교 정보통신공학과

{alsry4706*, nhwzzang, jwk5206, whdydwls02}@pukyong.ac.kr, {kdj6306, jlee0315}@pknu.ac.kr

Smart Monitoring System by Object Detection

Minkyoo Kang*, Hyeonwoo Nam, Jeongwan Kang, Yongjin Jo, Dongjae Kim and Jaewook Lee

Dept Information and Communications Eng., Pukyong National University

요약

본 연구에서는 CCTV 시스템의 저장 효율성과 정보 활용성을 극대화하기 위한 접근법을 제안한다. 객체 탐지 기술을 통해 탐지된 프레임은 원본 해상도로, 탐지되지 않은 프레임은 저해상도로 구분하여 저장 공간을 최적화하였다. 또한, GFPGAN 기반 초해상화 기법을 적용하여 원거리 객체나 확대된 얼굴도 선명하게 식별할 수 있도록 개선하였다. 더불어, 웹 서버를 구축하여 실시간 모니터링과 업스케일링 요청이 가능하도록 설계하였으며, 이를 통해 사용자는 원격지에서도 탐지 결과를 확인하고 식별을 진행할 수 있다. 본 시스템은 저장 용량 절감과 동시에 시각적 정보의 활용도를 높여, CCTV 모니터링의 효율성을 극대화한다.

I. 서론

현대 사회에서 CCTV는 보안과 안전 관리를 목적으로 가정, 상가, 공장, 사무실 등 다양한 장소에 널리 설치되고 있다. 일반적으로 1,080p 해상도의 카메라 4대를 30프레임으로 30일간 연속 저장하기 위해서는 약 6~7TB의 대용량 저장 공간이 필요하다. 해상도가 높아질수록 저장 용량 요구가 급격히 증가하며, 이를 해결하기 위해 화질을 낮추거나 저장 기간을 단축하는 방식이 사용되지만, 이에 따라 중요한 객체 식별이 어려워질 수 있다. 특히 저해상도나 노후화된 CCTV는 인식 실패가 빈번하며, 실제로 약 3대 중 1대는 사람의 얼굴을 명확히 식별하지 못한다는 통계도 존재한다.

본 연구에서는 이를 해결하기 위해 CCTV 영상 내 객체 탐지 여부에 따라 프레임을 구분하여 저장 효율을 극대화하는 방법을 제안한다. YOLO v8(You Only Look Once, Version 8) 모델을 활용하여 객체가 탐지된 프레임은 원본 해상도로 유지하고, 탐지되지 않은 프레임은 저해상도로 압축하여 저장 용량을 최적화한다.[1]

또한, 탐지된 객체 중 얼굴에 대해서는 GFPGAN(Generative Facial Prior Generative Adversarial Network)을 통해 초해상화(Up-scaling) 처리하여 식별 정확도를 높인다. 이러한 방식은 웹 서버를 통한 실시간 모니터링이 가능하며, 저장 공간 절감과 동시에 객체 인식의 신뢰성을 극대화할 수 있다.

II. YOLOv8 기반 객체 탐지를 통한 용량 확보

본 연구에서는 Open CV를 활용하여 CCTV 영상을 프레임 단위로 캡처한 뒤, YOLO v8 객체 탐지 모델을 적용하여 프레임 내 객체 유무를 실시간으로 판별한다. YOLO v8은 높은 프레임 처리 속도(FPS)와 우수한 객체 탐지 정확도를 제공하므로, 실시간 감시 환경에 적합하여 사용한다.[2][3]

이후 영상을 객체 탐지 결과에 따라 탐지된 프레임과 탐지되지 않은 프레임 두 가지 프레임으로 분류한다. 탐지된 프레임은 원본 해상도를 유지하여 중요한 정보를 손실 없이 저장하였으며, 탐지되지 않은 프레임은 저

해상도로 변환하여 용량을 최적화한다. 이를 따로 처리하는 이유는 해상도를 빈번하게 변경하면서 하나의 영상으로 병합할 경우, 동영상 압축 과정에서 효율성이 저하되어 저장 용량이 최적화되지 않고 저장 공간의 낭비가 발생하기 때문이다. 또한 객체가 탐지되지 않은 구간이라도 사건 발생 여부를 확인하거나, 보조적인 정보 제공에 필요한 경우가 존재하므로, 데이터를 완전히 삭제하지 않고 최소한의 화질로 기록을 유지할 필요가



[그림 1] 사람의 유무에 따른 프레임 구분

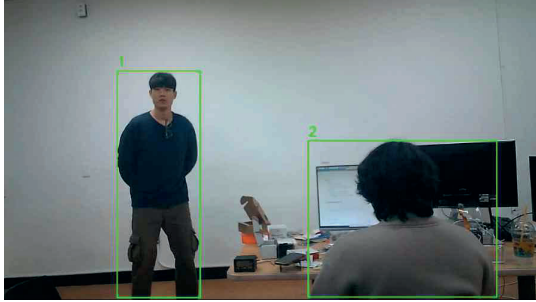
있다. 이를 기반으로 [그림 1]에 나타난 것처럼 왼쪽의 탐지된 프레임과 오른쪽의 탐지되지 않은 프레임을 구분하여 각각 다른 처리 방식을 적용하였다. 이렇게 프레임을 구분한 후, FFMPEG를 이용하여 각 프레임의 해상도 조정과 압축을 진행하였다. FFMPEG는 crf (Constant Rate Factor) 파라미터를 통해 가변 비트레이트 기반의 화질 조정을 지원하며, 이 값은 0에서 51 사이로 설정할 수 있다. 값이 낮을수록 화질이 높아지며, 값이 높을수록 압축률이 증가한다.

본 연구에서는 libx264 코덱을 사용하였고, 탐지된 프레임에 대해서는 crf 값을 18로 설정하여 무손실에 가까운 화질을 유지하였다. 반면, 탐지되지 않은 프레임에 대해서는 crf 값을 28로 설정하여 상대적으로 높은 crf 값을 적용하였다. 이러한 방식으로, 객체가 존재하는 중요한 프레임은 고해상도로 보존하면서도, 객체가 없는 구간의 저장 용량을 대폭 절감할 수 있다.

III. GFPGAN 기반 화질 개선

CCTV 영상에서 객체가 탐지되더라도, 대상이 카메라와 멀리 떨어져 있거나 얼굴이 작게 촬영된 경우, 원본 해상도(1,080p)로는 식별이 어려운

문제가 발생한다.



[그림 2] 촬영 영상 중 일부 장면

[그림 2]에 나타난 Person 1의 사례를 보면, 1,080p 해상도로 촬영되었음에도 불구하고 거리가 멀어 얼굴 식별이 명확하지 않다.

이를 해결하기 위해, 본 연구에서는 AI 기반 Super Resolution 기술인 GFPGAN을 활용하여 객체의 얼굴을 고해상도로 복원하는 시스템을 제안한다. 사용자는 식별이 필요한 장면을 [그림 2]와 같이 캡처한 뒤, 대상 객체를 선택한다.



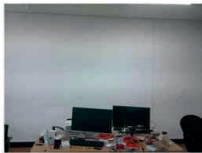
[그림 3] 객체 선택

[그림 4] 얼굴 확대

객체가 선택되면 [그림 3]과 같이 해당 프레임을 추출하고, YOLOv8 모델을 통해 얼굴 영역을 탐지한다. 탐지된 얼굴 부분은 별도로 분리되어 후속 처리에 사용된다. 추출된 얼굴 영역은 [그림 4]와 같이 GFPGAN을 통해 업스케일링 과정을 거친다. 이 과정에서 해상도가 향상되어 원본 영상에서는 식별이 어려웠던 대상의 얼굴을 명확하게 볼 수 있게 된다. 본 연구에서 제안한 시스템은 사용자가 필요한 순간에 원하는 객체를 확대 및 복원할 수 있어, CCTV 영상을 통한 객체 식별의 신뢰성을 크게 높인다.[4]

IV. 모니터링 시스템 구축

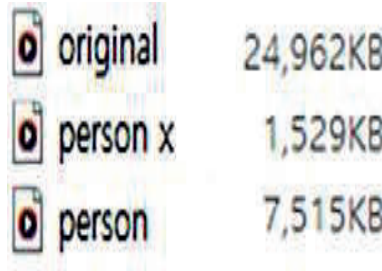
2025-01 2025-02 2025-03 2025-04 2025-05 2025-06 2025-07 2025-08 2025-09 2025-10 2025-11 2025-12 - 8월 22일



[그림 5] 모니터링 웹 서버

[그림 5]와 같이 CCTV를 실시간으로 모니터링할 수 있고, 월별로 영상을 객체 탐지를 통한 저장과 업스케일링 등을 할 수 있도록 모니터링 서버를 구현하였다. [그림 5] 상단의 해당 월을 클릭하면 [그림 6]과 같이 해당 월의 영상을 볼 수 있으며, 이 영상을 객체 탐지를 통해 프레임을 구분하여 서버에 저장한다. 이후 필요한 상황에서는 해당 영상의 한 장면을 업스케일링 하여 필요한 부분의 얼굴을 고해상도로 복원할 수도 있다. 이 웹 서버를 통해 실시간으로 모니터링하면서 동시에 객체 탐지 및 업스케일링을 처리함으로써 효율적인 모니터링이 가능하다.

V. 결과



[그림 7] 저장 용량 비교



[그림 8] 업스케일링 후 결과

[그림 7]은 실험을 통해 촬영된 영상의 저장 용량 비교 결과를 나타낸다. 실험에서는 객체가 탐지된 구간과 탐지되지 않은 구간을 각각 별도로 저장하였으며, 비교를 위해 원본 영상도 함께 첨부하였다.

실험 결과, 원본 영상과 비교하여 약 37%의 용량이 절감되었음을 확인할 수 있었다. 이는 객체 탐지 여부에 따라 해상도를 조정하고 효율적으로 압축한 결과로, CCTV 시스템의 저장 공간을 효과적으로 절약할 수 있음을 의미한다.

[그림 8]은 업스케일링 처리 후의 결과를 나타낸다. 복원된 얼굴 영상에서는 눈, 코, 입 등의 이목구비가 선명하게 드러나며, 식별이 용이해진다. 이러한 결과는 기존 1,080p 영상에서 확인하기 어려웠던 세부 정보를 선명하게 복원하여, CCTV 영상 분석 시 중요한 증거 자료로 활용될 수 있음을 보여준다.

VI. 결론

본 논문에서는 객체가 탐지된 프레임은 원본 해상도로 저장하여 모니터링 및 식별의 정확성을 유지하였으며, 탐지되지 않은 프레임은 해상도를 낮춰 저장 공간을 최적화한다. 또한, GFPGAN을 활용한 초해상화 기법을 적용하여, 원거리 객체나 확대된 얼굴 영역에서도 선명한 식별이 가능하다.

이와 더불어, 웹 서버를 구축하여 실시간 모니터링이 가능하고, 사용자의 업스케일링 요청을 통해 초해상화를 처리할 수 있다. 이를 통해 원격지에서도 객체 탐지 결과를 확인하고, 필요시 특정 객체의 해상도 개선을 즉시 수행할 수 있다. 이러한 구조는 시스템의 처리 속도를 높이고, CCTV 영상의 저장 효율성과 활용성을 극대화할 수 있다.

본 시스템은 저장 공간의 절약과 객체 식별 능력 향상뿐만 아니라, 사용자 접근성을 높여 실시간 보안 관리의 효율성을 극대화할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] 김성익. (2023). "YOLO v3를 이용한 CCTV 영상파일 저장공간 확보 모델". 한국컴퓨터정보학회논문지, 28(1), 65-70.
- [2] Yong-Hwan Lee, Youngseop Kim, "Comparison of CNN and YOLO for Object Detection", in Journal of the Semiconductor & Display Technology, Vol. 19, No. 1. March 2020.
- [3] Joon-Yong Kim, "A comparative study on the characteristics of each version of object detection model YOLO" in 2023 Summer Conference of the Korea Computer Information Society Vol. 31, No. 2, Nov, 2023
- [4] Vijul Shah, Brian B. Moser, Ko Watanabe, Andreas Dengel, "Webcam-based Pupil Diameter Prediction Benefits from Upscaling" in German Research Center for Artificial Intelligence (DFKI), Germany