

마스크 착용 유무를 탐지하는 지능형 CCTV 설계

Design of Mask Detection Intelligent CCTV

요 약

최근 전 세계적으로 코로나의 심각성이 완화되면서 일상 회복에 많은 관심을 보이고 있다. 그러나 해외에서 마스크 착용 의무화를 해제하면 다시 확진자가 급증하는 현상을 보이고 있기 때문에 원하는 일상으로 돌아가려면 감염을 줄이기 위해 실내 마스크 착용은 착실히 이행되어야 한다. 이를 효율적으로 감시하고 관리하기 위해 object-tracking 기능을 지닌 지능형 CCTV를 설계하여 사람이 항상 감시해야하는 수고를 덜고 엄격한 기준 하에 마스크 착용을 관리할 수 있도록 한다.

1. 서 론

최근 코로나 바이러스라는 전례 없는 전염병 사태를 겪으면서 사람 간의 접촉을 최소화하고 전염률을 최대한 낮추려고 노력 중이다. 마스크 착용은 감염병 위기 상황에서 사람들의 건강을 보호하기 위한 방역지침이다. 우리나라의 경우, 이를 위반하고 행정명령을 한 시설, 장소 등에서 마스크를 착용하지 않으면 과태료를 부과하고 있고 이를 감시하기 위한 신고, 관리 제도를 도입하고 있다. 비록 전 세계적으로 코로나의 위험성 낮아지면서 실외 마스크 착용 의무화를 완화하고 있는 추세이지만, 계속 되는 코로나 변이와 늘어나는 확진자로 인해 공항, 콘서트장, 대중교통 등 실내 밀집 구역의 마스크 착용은 여전히 엄격하게 관리되고 있다. 또한 미국, 유럽의 몇몇 국가들에서는 완화 후 다시 폭증하는 확진자로 다시 마스크 의무화를 적용한 사례도 존재한다.

다시 평범한 일상으로 돌아가기 위해서는 실내 밀집 구역의 코로나 확산을 막기 위해 마스크 착용은 필수가 되어야 하고 지속적인 감시가 필요하다. 이것을 사람이 수행하기에는 번거로움이 존재하므로 지능형 CCTV가 대신 수행할 수 있도록 한다. 항상 현장에서 감시하는 방법 대신, 마스크 착용 유무를 판단하는 지능형 CCTV를 사용하여 감시하는 역할을 대신할 수 있도록 한다. 해당 CCTV는 영상 압축 과정에서 발생하는 정보를 활용하여 객체를 탐지하고 적합한 복잡도의 연산을 활용하여 효율적이고 실시간 탐지가 가능할 수 있도록 설계한다.

2. 영상 압축 관련 연구

해당 절에서는 현재 사용되는 지능형 CCTV의 영상 처리 구조를 파악하여 어떤 방식으로 카메라와 영상을 제어할지 목표를 설정하고 각각의 과정에서 사용될 영상 압축 방식을 연구하여 적합한 연산을 직접 구현할 수 있도록 한다.

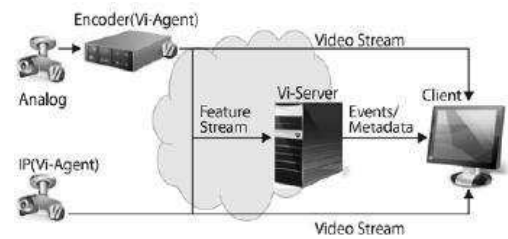
2.1. 지능형 CCTV 영상 분석 구조

[1] 지능형 CCTV 영상 분석 방법은 이벤트를 탐지하는 지점에 따라 에지 기반의 방법, 서버 기반의 방법으로 나뉜다.

- 에지 기반: 카메라 혹은 카메라와 연결된 영상 인코더에서 영상을 분석하는 방법이다. 영상을 전송하기 위한 압축 과정에서 발생하는 신호로 왜곡 없이 영상을 분석할 수 있는 장점이 있다. 그러나 장비의 연산 능력이 제한되기 때문에 복잡도가 높은 영상 분석 기법을 적용할 수 없고 추후 기능의 업그레이드가 어려운 문제점이 있다.

- 서버 기반: 카메라와 연결된 별도의 서버에서 영상을 전송받아서 지능형 영상 분석을 수행하는 집중형 구조이다. 카메라의 성능에 대한 의존성이 없기 때문에 기존 시스템에 적용하기 용이하지만, 사용되는 CCTV가 늘어날수록 서버가 처리할 수 있는 카메라가 제한되기 때문에 확장성에 대한 문제가 있다.

최근에는 에지 기반과 서버 기반의 단점을 보완하기 위해 하이브리드 기반 지능형 영상 분석 방법도 연구되었다.



(그림 1) IPoIP 구조 [2]

대표적으로 AgentVi사의 IPoIP가 있으며 카메라 서버에 특정 정보만을 전송하여 불필요한 정보를 전송하지 않는다. 또한 모든 프레임을 전송하지 않고 특정 임계값이 넘는 모션에 대해서만 전송을 하기 때문에 이를 통한 네트워크 데이터 사용률을 줄일 수 있다.

2.2. DCT + Quantization

영상 데이터 압축의 기본 원리는 인간 시각이 인지할

수 없는 정도의 범위 내에서 영상 정보를 제거하거나 줄이는 것이다. DCT는 영상 내의 데이터를 주파수 도메인으로 변환시키는 과정이다. 그 후 저주파에서 고주파 성분으로 정렬을 하고 Quantization 과정을 거쳐 불필요한 정보를 제거한다. 그 결과 인간의 눈으로는 인지하지 못하지만 영상의 데이터가 감소한 효과를 얻을 수 있다.

2.3. Motion Estimation/Compensation(ME/MC)

ME/MC는 연속된 프레임 간의 공통점을 찾아내어 유사한 블록에 대해서는 픽셀값이 아닌 모션 벡터로 표현하여 데이터 전송량을 줄이는 방법이다. 연속된 프레임에서 최대한 공통된 부분을 찾아내고 정확한 움직임을 찾아내는 것이 ME/MC의 핵심이라고 할 수 있다. ME/MC를 사용했던 비디오 코덱은 프레임 참조를 하지 않는 I-frame(Intra), 이전의 프레임을 참조하는 P-frame(Predictive), 이전 혹은 이후의 프레임을 모두 참조하는 B-frame(Bi-predictive)을 적절히 사용하여 오류를 최소화하면서 전송 데이터를 줄였다.

3. Object-tracking 관련 연구

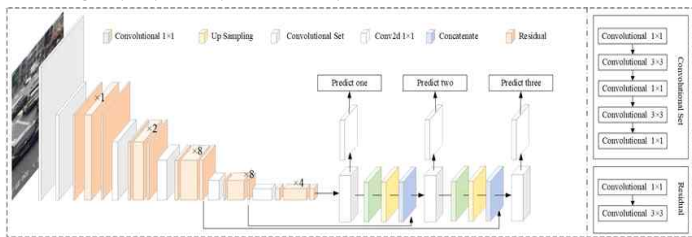
해당 절에서는 최근에 사용되는 동영상에서의 객체 검출 기술을 연구하고 지능형 CCTV의 실시간 탐지에 적합한 기술의 성능을 비교해가며 결정할 수 있도록 한다.

3.1. Convolution Neural Network(CNN)

CNN은 입력된 이미지로부터 위상 특성을 추출하는 피드-포워드 네트워크 방식을 취한다. 입력 이미지에서 특성을 제거한 다음, 분류기를 통해 추출된 특징을 분류한다. 따라서 Scaling, Squeezing, Rotation 등 기본 형상 변환에 강인한 특성을 갖는다[5]. 기존의 CNN보다 효율을 높이기 위해 R-CNN, SppNET, Fast R-CNN, Faster R-CNN 순으로 연구 개발되었다.

3.2. YOLO

YOLO는 이미지 내에 존재하는 객체와 해당객체의 위치는 이미지를 한번만 보고 예측할 수 있는 알고리즘이다. 입력 이미지를 CNN을 통해 Tensor의 그리드 형태로 나누고, 각 구간에 따라 경계 박스와 클래스 확률을 생성하고 해당 구역의 객체를 인식한다.



(그림-2) YOLO v3의 구조 [6]

3.3. CNN과 YOLO 비교

R-CNN은 전체 이미지를 유지하면서 후보 영역을 나중에 나누는 반면, YOLO는 CNN을 처리하기 전에 미리 이미지를 나눈다. 또한 YOLO v3는 정확도 향상을 위해 다

차원 클러스터를 생성하지만, Faster R-CNN은 지원하지 않는다.

YOLO는 Fast R-CNN에 비해 정확한 지역화에서 낮은 정확도를 보이지만 추론 시간과 훈련에 소요되는 시간이 짧다[7]. 또한 검출 성능에서 직접 대응하는 손실 함수에 대해 훈련하고 처리 시간 측면에서는 실시간 객체 검출이 가능하다. 따라서 여러 강점을 통해 지능형 CCTV에서는 YOLO를 사용하여 마스크 착용/미착용 객체를 탐지하는 것이 적합하다.

4. 마스크 착용 여부 판단 지능형 CCTV 설계

2절에서 조사했던 연구 결과, 연산 방식으로 마스크의 착용 유무를 탐지하는 지능형 CCTV를 설계해보겠다. Main Camera에서 얻은 영상을 압축하여 Sub-Camera가 Object-tracking을 처리하는 데이터의 양을 줄일 수 있도록 한다.

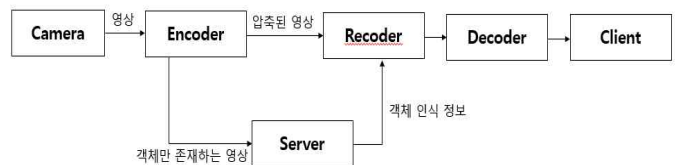
4.1. 고려 사항

현재 통상적으로 사용되는 CCTV는 화질이 640x480의 20~25 프레임부터 최대 1920x1080(FHD) 30프레임까지 사용되고 있다. 공공장소나 대중교통에서 사용되는 CCTV는 2018년 기준 95%가 50만 화소 미만의 저화질 CCTV로, 영상 복원 방법이나 고화질 CCTV로의 전환이 차츰 이루어지고 있음을 통해 CCTV도 고화질 영상 처리가 가능해야 한다.

목표로 설정한 지능형 CCTV는 최소 HD 화질을 처리할 수 있어야 하며 동시에 실시간으로 감시하기 위해선 압축, 복원에 소요되는 처리 시간이 짧아야 한다[3].

기존에 설치된 CCTV를 지능형 CCTV의 역할을 수행할 수 있도록 녹화하는 과정에서 Object-tracking 기능을 추가한다.

4.2. 구조



(그림-3) 지능형 CCTV 구조

• Camera: 고정된 Wide-viewed angle로 이벤트를 촬영한다. 감시 구역의 모든 정보를 촬영하는 카메라로 Client가 실시간 모니터링을 할 수 있도록 온전한 영상을 전송해야 한다. Object Tracking 효율을 늘리기 위해 전

송 과정에서 압축하는 절차를 거치고 해당 과정 중에서 발생하는 정보를 이용하여 Server로 전송한다. CCTV의 특성상 고정된 화면을 촬영한다는 점을 사용하여 영상 중에서 배경 영역을 제거하고 객체만 존재하는 영상을 Server로 전송하면 네트워크 부하를 줄이고 객체 탐지 처리 효율을 높일 수 있다.

- Server: 영상에서 배경 영역과 분리되어 객체들만 존재하는 프레임을 전송 받는다. 사람 객체만을 탐지하며 그 후 사람들의 마스크 착용 유무를 판단하는 object tracking 연산을 처리한다. 마스크 미착용 객체의 등장 이벤트 발생 시, 현장에서 경고 알람을 재생하거나 Client에게 영상 정보를 주어 즉각적인 조치를 할 수 있도록 관리한다.



(그림-4) 마스크 착용 탐지 예시

- Recoder: 카메라에서 촬영하는 영상 정보를 기록한다. 기존의 CCTV 시스템과 동일
- Client: CCTV에서 촬영하는 영상을 볼 수 있는 대상으로 규모가 큰 건물의 경우 통합 관제실에 해당된다. Camera에서 전송되는 일반적인 영상과 Server에서 처리한 마스크 미착용 대상의 모습도 확인할 수 있다.

5. 결론 및 향후 연구

본 논문에서는 기존의 CCTV에도 기능을 추가할 수 있는 마스크 탐지 지능형 CCTV 설계를 제안하였다. 영상 압축 과정 중의 정보 중 Yolo를 통하여 마스크를 착용한 객체와 착용하지 않은 객체를 탐지할 수 있다.

향후 연구로는 압축 효율을 높일 수 있는 방법과 효율적이고 정확한 object-tracking을 위한 영상 전달 알고리즘을 연구할 계획이다.

참고문헌

- [1] 정치윤, 한종욱. “지능형 영상분석 이벤트 탐지 기술 동향”, 전자통신동향분석, 27(4), 114-122, 2012
- [2] AgentVi, “Video Analytics Architectures”, 2012
- [3] 이상훈, 추성권, 권기범, 조남익. “지능형 CCTV 시스템을 위한 보행자 검출과 재식별 기술”, 한국통신학회지 (정보통신), 34(7), 40-47, 2017
- [4] 최승현, 김도현, 김형현, 김윤. “CNN을 이용한 재난 예경보 시스템”, 한국컴퓨터정보학회논문지, 24(2),

25-33, 2019

- [5] 이용환, 김영섭. “객체 검출을 위한 CNN과 YOLO 성능 비교 실험”, 반도체디스플레이기술학회지, 19(1), 85-92, 2020
- [6] Kwanghyun Kim, Sungjun Hong, Baehoon Choi, Euntai Kim. “Probabilistic Ship Detection and Classification using Deep Learning”, Applied Sciences, vol.8, no.6, 2018
- [7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-time Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 779-788, 2015.