**마스크 착용 유무를 감지/추적하는 지능형 CCTV**

2017103962 김동규

**1. 서론**

**1.1. 지능형 CCTV**

최근 몇년간 스마트시티 프로젝트 등이 추진되면서 공공기관이 운영하는 CCTV 대수는 더욱 늘어나고 있다. 또한 민간시장에서도 홈-security 등을 위해 소형 CCTV 디바이스 설치가 확대되는 추세다. 이처럼 CCTV가 늘어나면서 관제 인력이 부족해지는 문제가 발생하고 있다. CCTV 설치 대수가 기하급수적으로 늘어나면서 모든 영상을 사람이 실시간으로 감시하는 것은 불가능에 가까워졌다. 이에 따라 최근에는 영상 분석 기술을 활용한 지능형 CCTV 도입이 일반화되는 추세다.

지능형 CCTV는 실시간으로 영상을 분석해 움직임이 있는 물체를 감지·분류하고, 사전에 정의된 사건을 감지해 자동 식별하고 이를 감시자에게 알려 사람이 24시간 영상을 감시할 필요없이 선별적인 관제가 가능하다. 이는 실제로도 식별하는 것이 중요한 상황이나 행위를 놓치는 것을 방지하는 효과가 있어 실시간으로 위험 상황에 대한 대응이 가능하게 돕는다. 또한 선별적으로 유의미한 상황이 담긴 영상만을 저장하고 관리할 수 있기에 더 효율적이다.

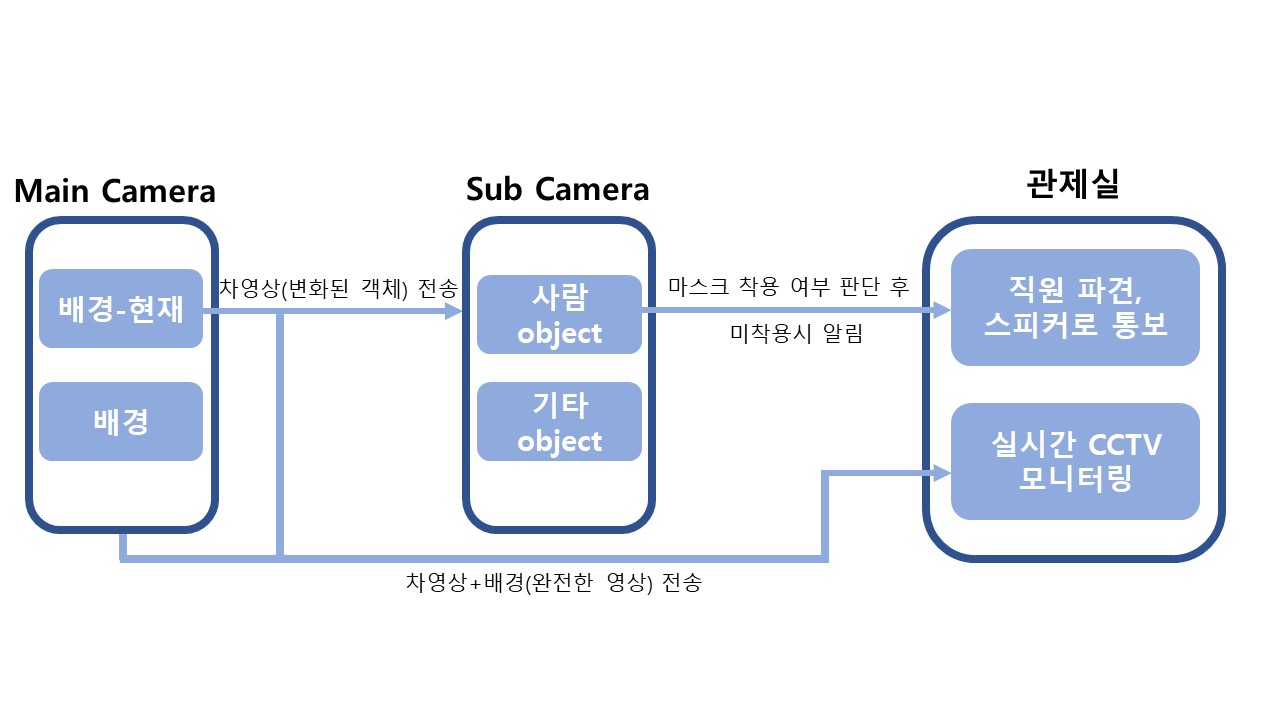
**1.2. 지능형 CCTV의 원리**

딥 러닝 기반의 '행동인식 기술', 행동인식 기술이란 인간의 신체 동작을 자동으로 인식하는 기술이다. 딥 러닝으로 영상 기반의 행동을 인식하기 위해서는 CNN(Convolution Neural Network) 기반의 방법과 LSTM(Long Short Term Memory) 기반의 방법이 있다. 이 두 방법은 모두 움직이는 모션 정보를 추출하기 위해 비디오 영상의 여러 프레임을 종합적으로 분석하는 데에 중점을 두는데, CNN 기반 방법은 연속적인 프레임을 쌓아 두고 3D 커널을 적용함으로써 움직임 정보를 잘 축적한 특징맵을 생성하고, LSTM 기반 방법은 영상을 프레임별로 분석하되 과거 프레임에 대한 기억을 Cell이라는 구조를 통해 조절한다. 즉, 이전 프레임이 행동인식에 있어 중요한 정보를 가지고 있다고 판단된다면 이 정보를 유지하고, 이전 프레임에 중요한 정보가 없다고 판단된다면 해당 프레임의 정보를 잊는다. 이렇게 CNN과 LSTM 방법 모두 연속된 비디오 프레임에서 정보를 분석하여 특징 벡터를 추출한 후 행동에 대한 스코어 값을 계산하여 행동을 식별한다.

**2. 프로젝트 내용**

**2.1. 프로젝트 목표**

동영상 압축 기능은 스마트폰 등 모든 IT 정보가전 제품군에 탑재된 매우 중요한 기능이며, 특히 CCTV 등의 Surveillance System에는 필수적인 기능이다. 지능형 CCTV에는, Moving Object Tracking, Synchronized Main and Sub PTZ-Camera System 등이 주요 필수 요소이며 Killer-Application으로서 부각되기 위해서는, 필히 Cost-effective하여야만 한다. 특히 Synchronized Main & Sub PTZ-Camera System에서, Main Camera는 Wide-viewed angle로 Event를 기록하고, Sub-Camera는 Main Camera에서 포착된 사물의 상세 영상을 Close-up해서 기록하는 방식이 필요하며, 이를 위해서 Object Tracking 기능이 필수적으로 필요하다. 본연구에서는 동영상 압축 과정에서 발생되는 정보를 이용하여 (별도의 Object Tracking 기능을 탑재하지 않거나 해당 기능을 최소화시킨 상태에서, Sub PTZ-Camera에 필요한 Object Tracking 기능을 제공하여 Cost-Down을 실현하고, 동영상 압축과정에서 발생되는 정보를 효율적이고 지능적으로 이용하여 우수한 성능을 지닌 Intelligent CCTV System을 구현한다.

**2.2. 구조**

(그림-1) 지능형 CCTV 구조

**2.3. 핵심 기능**

**-** Main Camera

**>** 건물 CCTV는 감시하는 배경이 실내에 고정되어 있으므로 배경은 변하지 않음을 활용

> (현재 영상) - (기존 배경)으로 차영상을 구하게 되면 움직임이 생긴 객체를 얻을 수 있다.

> Sub Camera에게는 차영상만, 모니터링이 필요한 관제실엔 완전한 영상을 보낸다.

- Sub Camera

> 전달받은 객체를 사람과 기타 필요 없는 객체로 구분한다.

> 사람 객체에서 마스크를 tracking하여 마스크 착용 여부를 판단한다.

> 미착용 적발 시 관제실에 알람을 보내 이벤트 발생을 알린다.

- 관제실

> Main Camera에서 오는 실시간 CCTV 영상을 모니터링 할 수 있다.

> Sub Camera에서 적발한 마스크 미착용 사람에게 통보, 직원 파견 등 조치를 취한다.

**3. 기능 구현 설계**

**3.1. Main Camera**

**-** 역할: 영상의 압축과 관제실에서 관측하는 Wide View 영상을 전송

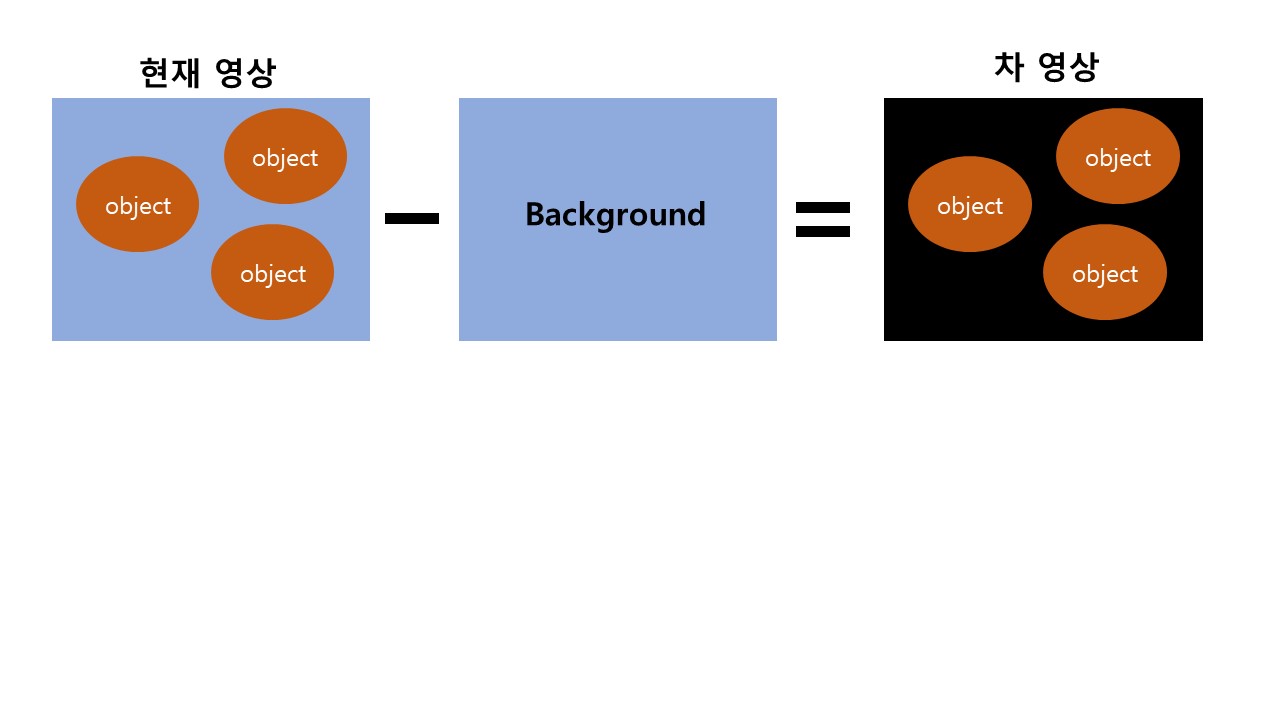
- 필요 도구: CCTV 역할을 할 고정된 카메라

> 카메라: 라즈베리파이 사용(여건이 안될 경우 웹캠 사용)

- 구현 방법: 차영상 기법 활용

> 기대 효과: Sub Camera가 모든 영상을 처리하는 것이 아닌 객체만 추출된 영상을 사용하여 Object Tracking의 효율을 증대시킨다.

> 객체 분할: 차영상 기법을 사용해 배경은 mask되고 객체만 있는 영상을 구한다



(그림-2) 차영상

**3.2. Sub Camera**

- 역할: Main Camera에서 전송 받은 영상에서 객체 검출, 추적, 상황인식

> 사람과 그 외의 객체를 검출

> 사람 객체에서 마스크를 검출하여 착용 여부 판단

> 마스크 착용을 하지 않은 사람을 추적, 관제실에 전송

- 목표로 정한 CCTV에서 Sub Camera가 줌인 하는 기능이 있지만 팀에서 활용할 수 있는 카메라로는 한계가 있을 경우 카메라 한 대로 Main, Sub Camera에서 처리하는 영상을 띄워서 과정을 확인

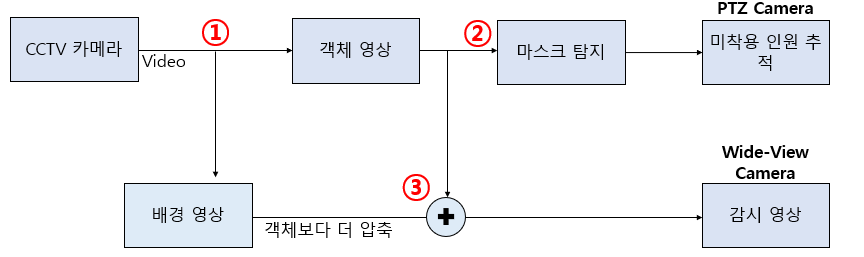
- 구현 방법: 특징점(feature)을 활용해 학습 알고리즘을 이용

**4. 프로젝트 최종 내용 및 결과**

**4.1. 프로젝트 설명**

Main Camera는 Wide-View로 촬영하면서 감시 구역을 계속 촬영하고 마스크 탐지 기능을 추가한다. 마스크 미착용 인원이 감지될 경우, PTZ Sub-Camera가 미착용 인원을 추적하여 사람의 상시 감시 없이 자동으로 대응하는 시스템이다.

**4.2. 프로젝트 상세 구조 설계**



(그림-3) 프로젝트 구조

① Main Camera에서 촬영한 영상을 객체만 존재하는 영상과 배경만 존재하는 영상으로 분리한다.

② 객체 영상에서는 움직이는 객체만 존재하고 이 객체의 대부분은 사람이 차지한다. 따라서 이 객체에 대해서만 마스크 탐지를 수행하고 미착용 인원은 PTZ Camera에게 전송하여 추적을 수행한다.

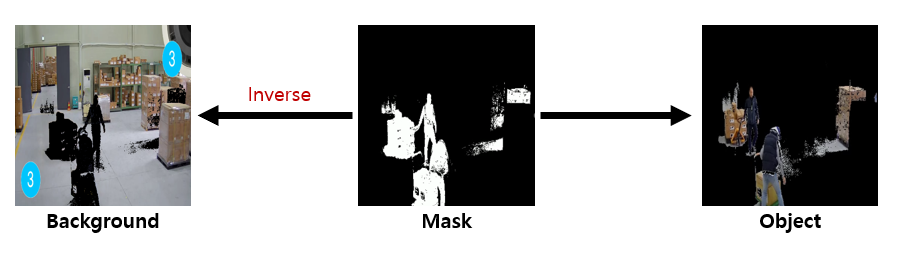
③ 분리했던 배경 영상은 비교적 객체 영상보다 덜 자세하게 표현해도 되므로 압축을 더 진행하고 객체 영상과 합하여 정상적으로 Wide-view angle로 볼 수 있는 CCTV 영상을 전송한다.

**4.3. 요구사항 및 구현방법**

**4.3.1. 배경과 객체 분리**

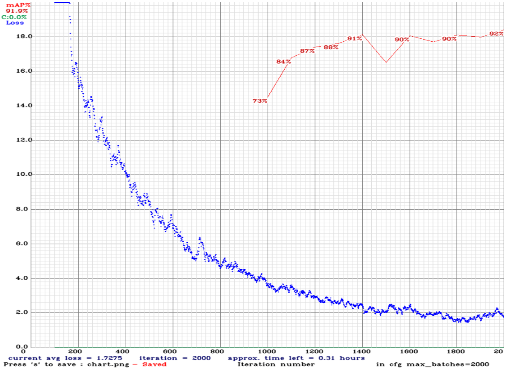
객체와 배경을 분리하기 위한 방법으로 과거의 프레임을 사용하도록 했다. 고정된 위치에서 촬영하는 CCTV의 특성을 활용하여 현재와 과거의 프레임을 비교했을 때, 움직임이 없는 부분은 배경, 움직임이 있는 부분을 객체라고 결정한 뒤 그에 맞는 마스크를 생성한다. 실외의 경우 먼지, 노이즈, 햇빛 등에 따라 변수가 많이 생겨 과거와의 차이로 배경과 객체 구분이 어려워지지만 마스크를 탐지할 CCTV인 만큼 실내에서 사용되기 때문에 영상의 노이즈에 대한 변수는 제거할 수 있었다.

마스크의 경우 객체 부분이 1, 배경 부분이 0으로 구성되어 있으며 배경 곧곧에 존재하는 화이트 노이즈를 제거하기 위해 ‘opencv’의 함수를 사용하여 배경과 배경 사이 혹은 객체의 내부에 존재하는 작은 노이즈를 제거하도록 했다. 이 마스크를 원본 영상에 곱하여 객체 부분은 원본과 동일하고 배경은 0으로 된 영상을 얻을 수 있다. 배경만 존재하는 영상의 경우 이 마스크를 반전시킨 뒤 적용하도록 한다.



(그림-4) 객체와 배경 분리

**4.3.2. 마스크 탐지**

 마스크 착용 인원과 미착용 인원을 선별하기 위해 마스크 탐지 기능을 수행할 학습 모델이 필요하다. 이를 본 프로젝트에서는 실시간 처리와 신뢰도가 중요하기 때문에 여러 학습 모델 중 ‘yolo’를 사용하기로 했다.

ㆍClass 구분

- mask: 마스크 착용

- improperly: 마스크를 턱에 걸친 상태(노마스크와 구분)

- no mask: 마스크 미착용

ㆍDataset (그림-5 yolo 학습 결과

|  |  |  |  |
| --- | --- | --- | --- |
| **Set** | **Images** | **with Mask** | **without Mask** |
| Training Set | 700 | 3047 | 868 |
| Validation Set | 100 | 278 | 49 |
| Test Set | 120 | 503 | 156 |
| Total | 920 | 3828 | 1073 |

ㆍ결과

|  |  |  |  |
| --- | --- | --- | --- |
| **모델명** | **Training Set** | **Validation Set** | **Test Set** |
| YOLO v4 | 99.65% | 88.38% | 93.95% |

여러 이미지로 충분히 학습시킨 결과, 임의의 영상에 적용해도 될 정도의 신뢰도를 얻을 수 있었고 이를 바로 프로젝트에 적용하도록 했다.

ㆍ적용 모습

대지, 실외, 하늘, 사람이(가) 표시된 사진

자동 생성된 설명

(그림-6) 예시 영상에 학습한 yolo 적용

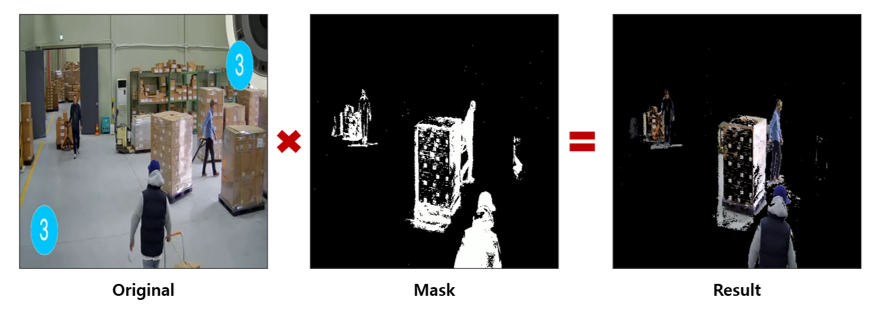
**4.3.3. 실시간 처리**

배경과 객체 분리, 압축, 마스크 탐지 과정 모두 복잡한 연산 과정을 거치면 높은 정확도를 얻을 수 있지만 그로 인해 처리 속도가 느려진다. 이 프로젝트는 저장된 영상이 아닌 CCTV에서 적용하므로 실시간 처리가 중요하다. 따라서 여러 함수를 적용하거나 값을 변경해가면서 최적의 결과를 얻을 수 있도록 한다.

**4.4. 동작 순서**

CCTV에서 영상을 촬영하고 사용자가 확인할 수 있는 영상으로 전환되는 과정을 차례대로 설명한다.

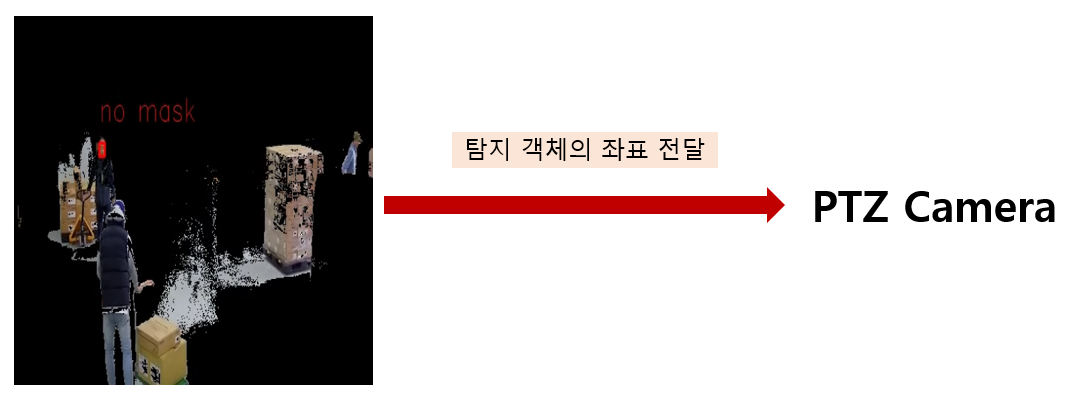
1) CCTV 영상에서 배경 제거



(그림-7) 원본 영상에서 배경 제거

촬영하고 있는 원본 영상에 5.3.1.절에서 언급한 마스크를 통해 배경을 제거하고 객체만 존재하는 영상으로 분리한다.

2) 마스크 탐지 및 미착용 인원 판별



(그림-8) 마스크 탐지 및 좌표 전달

객체만 존재하는 영상에서 학습시킨 yolo를 통해 mask detection을 수행한다. 이 때, 배경이 포함된 영상을 사용하게 되면 복잡해진 영상으로 인해 마스크 탐지를 올바르게 수행하지 않을 확률이 높기 때문에 객체만 존재하는 영상을 사용했다. 또한, 많은 연산이 필요한 탐지 과정에서 영상의 용량을 줄여야 하기 때문에 배경을 제거했다.

만약 마스크 미착용 인원이 발견되면 해당 좌표를 Sub-Camera(PTZ Camera)에게 전송한다.

3) Sub-Camera(PTZ Camera)에서 미착용 인원 추적

텍스트이(가) 표시된 사진

자동 생성된 설명

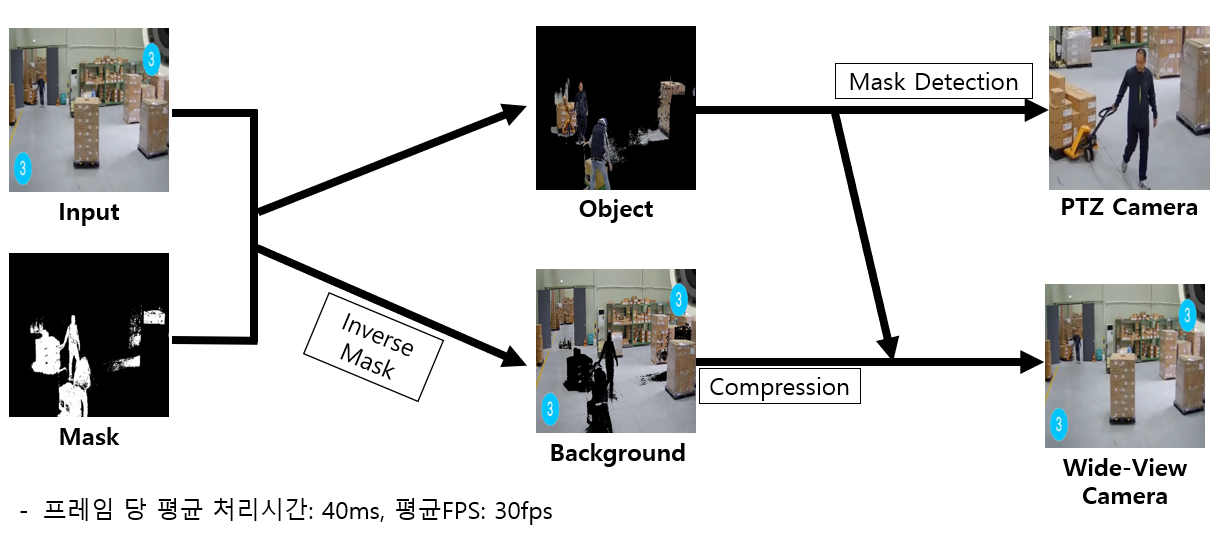
(그림-9) 미착용 줌인 및 인원 추적

미착용 인원에 대한 좌표를 전송받으면 Zoom-in과 Tracking 기능을 지닌 Sub-Camera가 이 인원에 대한 추적을 시작한다. 그리고 이 확대된 영상을 사용자에게 송출하여 걸맞은 대응을 수행한다.

4) Wide-View angle 영상 송출

3)의 과정과 동시에 이전 과정에서 얻은 객체만 존재하는 영상과 압축을 더 진행한 배경 영상을 합쳐 평범한 CCTV 영상을 얻는다. 이를 사용자에게 송출하면 정상적인 Wide-view angle 영상을 추적을 하고 있는 Sub-Camera 영상과 함께 볼 수 있다.

**4.5. 프로젝트 최종 구조**



(그림-9) 프로젝트의 동작 순서 및 구조

(그림-8)은 앞서 설명했던 프로젝트의 처리 순서 및 구조다. 결과적으로 사용자는 Wide-View 영상, PTZ Camera에서 얻은 추적 영상을 동시에 확인할 수 있고 이는 항상 Wide-view 영상을 감시하고 있어야 하는 고생을 덜 수 있다. 한 프레임 당 이 모든 연산 과정을 거치는데 평균 40ms의 처리 시간이 소요되고 영상에서의 평균 fps는 30fps로 충분히 실시간 영상에도 활용 가능함을 알 수 있다.

**4.6. 결론 및 개선 사항**

**4.6.1. 결론**

본 프로젝트를 통해 특정 유형의 객체를 감지하고 추적하는 지능형 CCTV를 구현해보았다. 예시로 제시한 영상 이외에도 여러 CCTV 영상에 적용한 결과 모두 정상적으로 기능을 수행했다. 이를 통해 CCTV 영상으로 감시할 수 있는 동시에 추적하는 영상을 통해 재빠른 대처가 가능하다. 또한 서버 데이터베이스와 연동할 경우, 어느 곳이 감시에 취약한지 알 수 있고 더 나아가 다른 학습 모델을 적용시키면 마스크 이외에도 다른 유형의 탐지를 진행할 수 있을 것이다.

**4.6.2. 개선 사항**

ㆍPTZ Camera가 tracking을 수행하는 만큼 카메라의 개수에 따라 추적할 수 있는 인원이 한정될 것이다. 이번 프로젝트에서 구현한 것처럼 인원 한 명을 추적하는 것 대신, 각 PTZ Camera 마다 구역(3x3, 5x5 등)을 나눠서 맡아 해당 구역에 미착용 인원이 감지되면 줌 인하는 형식으로 구현할 수도 있다.

ㆍTracking을 하는 만큼 마스크 탐지의 정확도를 꾸준히 향상시켜야 한다. 실시간 처리에 영향을 주지 않을 만큼의 정확도 향상이 필요하다.

ㆍ객체와 배경 분리가 더 깔끔하게 이루어져야 한다. 현재 프로젝트의 경우 한 장소에서 가만히 서 있는 인원은 정지되어 있으므로 객체가 아닌 배경으로 판별하는 문제가 있다. 이를 해결할 새로운 방법을 찾아야 객체와 배경의 분리가 더 확실하게 이루어질 것이다.

**4.6.3. 기대효과**

지능형 CCTV를 도입하면 적은 인원으로도 많은 CCTV를 동시에 감시할 수 있어 업무효율이 크게 증가한다. 모든 영상을 실시간으로 감시하기보다 확인이 필요한 장면만을 알려줌으로써 업무 부담을 줄이는 효과도 있다. 또한 중요한 영상만을 저장하는 것으로 데이터 용량 부족의 문제도 해결할 수 있다. 오작동이나 허위 알람도 줄어든다.