

# 다중 카메라 네트워크 가상의 관심선(Line of Interest)을 활용한 건물 내 재실자 인원 계수 방법론 개발

천희경<sup>ID\*</sup>, 박찬혁<sup>ID\*\*</sup>, 지식호<sup>ID\*\*\*</sup>, 노명일<sup>ID\*\*\*\*</sup>, Susilawati, Connie<sup>ID\*\*\*\*\*</sup>

Chun, Hwikyung<sup>ID\*</sup>, Park, Chanhyuk<sup>ID\*\*</sup>, Chi, Seokho<sup>ID\*\*\*</sup>, Roh, Myungil<sup>ID\*\*\*\*</sup>, Susilawati, Connie<sup>ID\*\*\*\*\*</sup>

## Developing an Occupants Count Methodology in Buildings Using Virtual Lines of Interest in a Multi-Camera Network

### ABSTRACT

In the event of a disaster occurring within a building, the prompt and efficient evacuation and rescue of occupants within the building becomes the foremost priority to minimize casualties. For the purpose of such rescue operations, it is essential to ascertain the distribution of individuals within the building. Nevertheless, there is a primary dependence on accounts provided by pertinent individuals like building proprietors or security staff, alongside fundamental data encompassing floor dimensions and maximum capacity. Consequently, accurate determination of the number of occupants within the building holds paramount significance in reducing uncertainties at the site and facilitating effective rescue activities during the golden hour. This research introduces a methodology employing computer vision algorithms to count the number of occupants within distinct building locations based on images captured by installed multiple CCTV cameras. The counting methodology consists of three stages: (1) establishing virtual Lines of Interest (LOI) for each camera to construct a multi-camera network environment, (2) detecting and tracking people within the monitoring area using deep learning, and (3) aggregating counts across the multi-camera network. The proposed methodology was validated through experiments conducted in a five-story building with the average accuracy of 89.9% and the average MAE of 0.178 and RMSE of 0.339, and the advantages of using multiple cameras for occupant counting were explained. This paper showed the potential of the proposed methodology for more effective and timely disaster management through common surveillance systems by providing prompt occupancy information.

**Keywords :** People counting, Multi-camera, Computer vision, DeepSORT, YOLOv4

### 초 록

건물에서 재난이 발생할 경우, 건물 내 인원을 신속히 구조하여 사상자를 최소화하는 것은 단연 최우선순위가 된다. 이러한 구조활동을 위해서는 건물 내 어디에 몇 명이 있는지를 알아야 하는데, 실시간으로 알기가 어렵다 보니 주로 건물주나 경비원 등 관계자의 진술이나 층별면적, 수용 인원과 같은 기초자료에 의존하는 실정이다. 따라서 빠르고 정확하게 재실 인원 정보를 파악하여 현장에 대한 불확실성을 낮추고 끝없는 내 효율적인 구조활동을 지원하는 것이 반드시 필요하다. 본 연구는 컴퓨터 비전 알고리즘을 활용하여 이미 건물에 설치되어 있는 여러 대의 CCTV가 촬영한 이미지로부터 건물 위치별 재실인원을 계수하는 방법론을 제시한다. 계수 방법론은 (1) 카메라별 관심선(LOI) 설정을 통한 다중 카메라

\* 정회원 · 네바다 주립대학교 박사후 연구원 (University of Nevada, Las Vegas · [hwikyung.chun@unlv.edu](mailto:hwikyung.chun@unlv.edu))

\*\* 정회원 · 아이티엔건축사사무소 R&D연구소 연구원 (ITM Engineers and Architects · [chpark133@gmail.com](mailto:chpark133@gmail.com))

\*\*\* 종신회원 · 교신저자 · 서울대학교 건설환경공학부 교수 (Corresponding Author · Seoul National University · [shchi@snu.ac.kr](mailto:shchi@snu.ac.kr))

\*\*\*\* 서울대학교 조선해양공학과 교수 (Seoul National University · [miroh@snu.ac.kr](mailto:miroh@snu.ac.kr))

\*\*\*\*\* 호주 Queensland University of Technology, School of Economics and Finance 부교수

(Queensland University of Technology · [c.susilawati@qut.edu.au](mailto:c.susilawati@qut.edu.au))

Received July 26, 2023/ revised August 21, 2023/ accepted August 28, 2023

라 네트워크 환경 구축, (2) 딥러닝을 활용한 모니터링 구역 내 사람 탐지 및 추적, (3) 다중 카메라 네트워크 환경을 고려한 인원 합산 세 단계로 구성된다. 제안된 방법론은 5층 건물을 대상으로 세 개의 시간대별로 수행된 현장 실험을 통해 검증되었다. 최종 결과는 89.9%의 정확도로 재실자를 인식하는 것으로 나타났으며, 층별, 구역별 합산 결과도 93.1%, 93.3%의 정확도로 우수했다. 층별 평균 MAE와 RMSE는 각각 0.178과 0.339이었다. 이처럼 실시간으로 제공하는 건물 내 재실자 정보는 초기 재난대응 단계에 신속하고 정확한 구조활동을 지원할 수 있다.

검색어 : 인원계수, 다중 카메라, 영상분석, DeepSORT, YOLOv4

## 1. 서론

재실 인원 정보를 빠르고 정확하게 취득하는 수단으로 최근에는 컴퓨터사이언스 분야의 객체 인식 및 추적 알고리즘이 비약적으로 발전하고 있다. 특히, 이를 건물에 기본적으로 설치되어 있는 CCTV 이미지에 적용하여 건물 내 인원을 계수하는 연구가 많이 이루어지고 있다(Sun et al., 2020). 그러나 건물 내 구역별 인원을 정확하게 계수하기 위해서는 여러 대의 카메라로 수집된 정보를 활용해야 함에도 불구하고, 많은 연구들이 단일 카메라를 활용한 재실자 인식 성능 개선에 집중하고 있는 실정이다(Dollar et al., 2011; Wen et al., 2017). 다수의 카메라를 인원 계수에 활용하는 연구도 소수 존재하지만, 같은 사람이 여러 대의 카메라에 동시에 등장할 때 정확하게 계수하는 것이 어렵고, 카메라 네트워크의 사각지대에 사람이 존재할 경우 계수를 놓치는 등 현실적인 한계가 있다(Wang 2013; Maddalena et al., 2014; Denman et al., 2015). 따라서, 여러 대의 카메라로 구성된 네트워크를 종합적으로 분석하여 건물 모든 구역에 존재하는 재실자의 위치와 수를 실시간으로 정확하게 파악할 수 있는 방법론이 필요하다.

이러한 방법론을 개발하기 위해서는 크게 세 가지 사항을 고려해야 한다. 첫째, 카메라 네트워크를 잘 설명해야 한다. 건물 내 카메라들의 위치를 정확하게 알아야 하고 각 카메라가 모니터링하는 영역을 잘 정의해야 한다. 이를 통해 카메라 간 관계를 설명하여 기존 연구가 어려웠던 겹치는 영역과 사각지대 등을 파악할 수 있다(Jave et al., 2008; Zhang et al., 2015). 둘째, 여러 대의 카메라가 각각 계수한 인원 정보를 어떻게 합산할 것인지 고민해야 한다. 기존 연구는 모든 카메라의 계수 정보를 단순히 합산하는 경우가 많아 중복 계수를 피하기 어려웠고, 중복 계수를 피하기 위한 연구들도 사전에 지정한 관심 영역만을 모니터링하게 하여 다수의 사각지대가 존재할 수밖에 없었다(Liu et al., 2016). 대형 건물이 많아지고 건물의 레이아웃도 복잡해짐에 따라 설치하는 카메라의 대수도 많아지고 있어 이에 적용 가능한 합산 방법이 필요하다. 마지막으로 합산한 정보를 어떤 단위로 제공할 것인지 고려해야 한다. 재난상황 시 효율적인 구조활동을 위해서는 최소 층이나 방 단위별 인원 정보를 제공해야 하지만, 기존 연구들은 건물 전체의 인원 정보만을 제공하는 경우가 많고 이 역시도 출입구

가 많은 경우 들어오는 사람과 나가는 사람의 수를 정확하게 파악하기 어렵다(Melfi et al., 2011).

따라서 본 연구는 컴퓨터 비전 알고리즘을 활용하여 이미 건물에 설치되어 있는 다중 CCTV 카메라 환경에서 건물 위치별 재실인원을 실시간 계수하는 방법론을 제시한다. 실제 건물에서의 실험을 통해 개발한 방법론의 현장 적용 가능성을 검증하고 여러 대의 카메라를 활용하여 인원 계수를 할 경우의 장점을 설명한다.

## 2. 문헌고찰

컴퓨터 비전을 활용하여 현장의 인원 수를 추정하는 연구로 다양한 방법들이 제안되었다. 예를 들어, Perng et al.(2016)은 대중교통 탑승자를 대상으로 머리 위에 설치된 오버헤드 카메라를 통해 이용자 수를 추정했다. Chae et al.(2020)은 건물 출입구에 설치된 카메라 영상으로부터 출입 인원을 실시간으로 계수하였고, Jung et al.(2018)은 철도역을 대상으로 보행자를 모니터링하여 테러, 화재 등 갑작스러운 재난이 발생했을 때 파란 동선을 최적화하는 연구를 수행했다. 특히, 다수의 이용자가 사용하는 시설물의 경우 화재 등 재난 발생이 대규모 인명피해로 이어질 수 있기 때문에 현장상황을 모니터링 할 수 있는 방재시스템의 일환으로 인원 수를 추정하는 방법이 제안되고 있다(Sharma et al., 2018). 이때 주로 사용된 기술은 단일 카메라 추적(Single-camera tracking, SCT) 기술로, 동영상 스트림 내 움직이는 사람의 위치를 인식하여 시간에 따라 추적하는 방식이다(Kalake et al., 2021). 객체를 인식(object detection)하는 알고리즘으로는 YOLO(You Only Look Once), Faster R-CNN(Region-based Convolutional Neural Network), SSD(Single-shot detector) 등으로 속도와 정확도를 모두 고려할 수 있다. 객체 추적(object tracking)은 이미지 프레임 내 객체의 크기, 위치, 색, 모양, 윤곽선 등 특징적인 정보 간의 유사도를 이용하여 객체의 변화를 추적하게 되는데 SORT(Simple Online Real-time Tracker), MIL(Multiple Instance Learning), KCF(Kernerlized Correlation Filters) 등의 알고리즘이 대표적이다.

기존의 인원 수를 세는 대부분의 연구는 단일 카메라 환경에 의존하고 있다. 개별 카메라의 상대적 위치와 촬영 영역을 일일이

고려하여 합산하는 것이 까다롭기 때문이다. 따라서 제한된 영역을 대상으로 인식과 추적을 수행될 수밖에 없었는데, 이를 해결하기 위해 카메라가 비추는 영역(Field of View, FOV)을 넓게 설정하게 되면 검출할 객체의 이미지가 너무 작아져 성능이 떨어지기 때문이다. 당연히 단일 화면에 의존하기 때문에 폐색(occlusion)과 시야확보(visibility)의 한계도 존재한다. 이러한 문제점을 해결하기 위해 연구자들은 다중 카메라를 활용하는 방안을 시도하였다. 예를 들어, Yu et al.(2013)은 병원 건물을 대상으로 각 카메라로 수집한 인원 정보를 도면에 표시하는 방식으로 인원을 계수하였다. 또한, Park and Chi(2020)는 건물 출입구만 대상으로 하던 계수 방법론에서 한층 나아가 층별로 위치한 사람 수를 파악할 수 있는 방법론을 제안하였다. 하지만 여전히 겹치는 영역이 없는 카메라 환경만을 대상으로 하는 등 현실적인 인원 계수에 적용하기에는 제한적인 수준에 머물고 있는 실정이다. 따라서 본 연구는 카메라 간 관계의 복잡성을 정의하고 개별 카메라로부터 수집되는 정보를 합산하는 인원 계수 모델을 제안한다.

### 3. 다중 카메라 환경에서의 인원 계수 모델 개발

본 연구에서 제시하는 인원 계수 모델은 (1) 카메라별 관심선(Line of Interest, LOI) 설정을 통한 다중 카메라 네트워크 환경 구축, (2) 딥러닝을 활용한 모니터링 구역 내 사람 탐지 및 추적, (3) 다중 카메라 네트워크 환경을 고려한 인원 합산 세 단계로 구성되며(Fig. 1), 단계별 세부내용은 다음과 같다.

#### 3.1 카메라별 LOI 설정을 통한 다중 카메라 네트워크 환경 구축

본 단계의 목적은 인원을 계수하기에 앞서, 개별 카메라들의 위치와 카메라별 LOI를 설정하여 구간별 이동 양상을 파악할 수

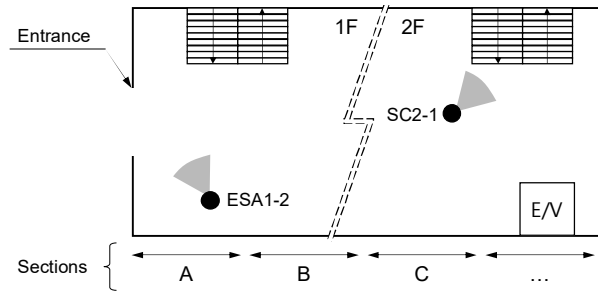


Fig. 2. Example of Camera Code System

있는 환경을 구축하는 것이다. 먼저 카메라의 위계는 설치 위치에 따라 (1) 건물 출입구(Entrance, E), 계단(Stair, S), 승강기(Lift, L), 복도(Hallway, H), (2) 층별 구역(각 층 좌측으로부터 A, B, C 순으로 증가), (3) 층수, (4) 카메라 번호(각 층 좌측으로부터 1, 2, 3 순으로 증가)로 구분하여 코드를 작성한다. 예를 들어, 출입구(E)와 계단(S)을 동시에 비추는 1층 A구역 2번째 카메라는 코드 ESA1-2를 부여받고, 계단(S)을 비추는 2층 C구역 1번째 카메라는 코드 SC2-1을 부여받는다(Fig. 2).

그 후 카메라별로 설정된 LOI를 통과하는 사람에게 ID를 부여하고 이를 각 카메라 코드에 매칭하여 각 구역별로 인원을 계수한다. LOI를 특정 카메라가 비추는 지역을 대표하여 설정함으로써 한 명의 사람이 여러 대의 카메라로부터 중복 카운팅 되는 것을 방지할 수 있고 사각지역도 최소화 할 수 있다. LOI는 수직, 수평, 대각선 혹은 이러한 선들의 혼합으로 설정할 수 있다(Fig. 3) 수직 방향의 LOI는 좌우로 이동하는 사람을 인식할 수 있고, 수평 방향의 LOI는 상하로 이동하는 사람을 인식할 수 있다. 큰 홀과 같이 대각선으로 가로질러 이동하는 사람을 인식하기 위해서는 대각선의 LOI가 필요하며, 모든 움직임이 가능한 구역에서는 여러 LOI를 혼합하는 것이 필요하다. 이때, 수직 및 수평방향 LOI는 주로 구역 구분에

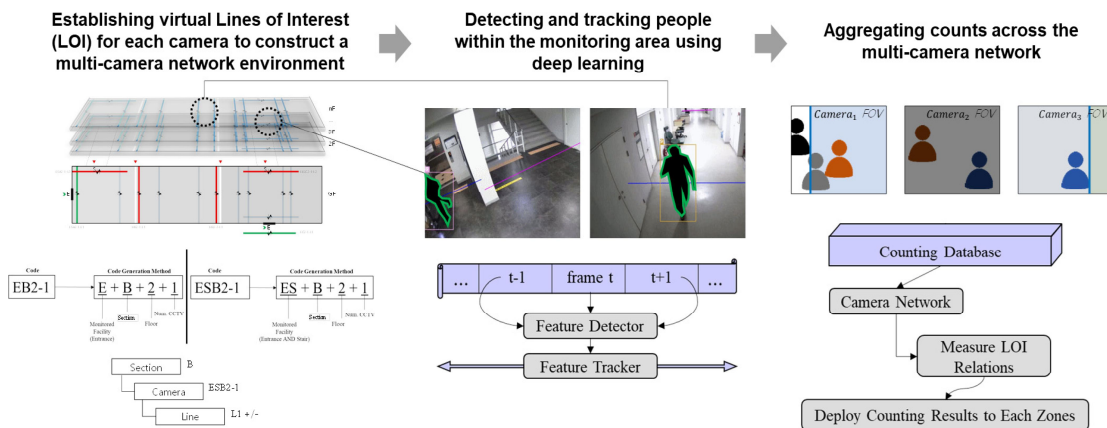


Fig. 1. Research Framework

따라 설정한다. 예를 들어 사람이 구역 A에서 구역 B로 이동하는 경우 구역 A를 대표하는 LOI를 통과하게 되면 구역 A의 인원은 +1이 되고, 그 후 구역 B를 대표하는 LOI를 통과하면 구역 B의 인원은 +1이 되는 반면 구역 A의 인원은 -1이 되게 된다. 본 연구에서는 건물 도면을 활용하여 수평과 수직 LOI를 기본적으로 설정하였다. 도면이 없는 경우, 설정한 LOI에 의해서 건물 층별 구역이 구분될 수 있다. LOI가 카메라 시야에서 너무 멀리 설정될 경우, 객체 인식 성능이 떨어질 수 있으므로 LOI를 조금 더 가깝게 보정할 필요가 있다. 이때 변경된 LOI에 맞춰 구역정보도 함께 업데이트 할 필요가 있다.

### 3.2 딥러닝을 활용한 모니터링 구역 내 사람 탐지 및 추적

본 단계는 영상에 나타난 사람을 탐지하고 추적하는 모델을 YOLOv4와 DeepSORT 객체 탐지 및 추적 알고리즘을 활용하여 개발한다. 객체탐지는 딥러닝을 통해 이미지 프레임에서 관심 객체(사람, 자동차 등)를 배경과 구분해 식별하는 컴퓨터 비전 기술이다. 본 연구에서는 건물 내 설치된 CCTV 카메라로부터 수집한 영상을 YOLOv4 객체 탐지기에 적용하여 탐지된 객체에 대응하는 바운딩 박스(Bounding Box)를 획득하였다. YOLOv4는 CNN을 사용해 객체를 탐지하는 알고리즘으로 특히 보행자에 대해 매우 빠르고 정확한 인식 성능을 보인다(Bochkovskiy et al., 2020). 객체추적은 탐지가 된 객체의 움직임과 경로를 추적하는 기술로, 객체 이동 경로의 유사도를 프레임 별로 계산하여 동일 객체 여부를 판단함으로써 객체를 추적한다. 본 연구에서는 DeepSORT 모델을 활용하여 탐지된 사람의 이동 경로를 도출하고, 이를 통해 건물 출입 정보와 건물 내 동선을 파악하였다. DeepSORT는 가장 널리 사용되는 객체 추적 아키텍처 중 하나로, 기존 SORT가 가진 한계(Occlusion, ID switch 등)를 개선한 성능을 보여주는 추적기이다. YOLO 알고리즘과 함께 사용하면 오랜 시간 동안 여러 객체를 추적할 수 있다(Wojke et al., 2017). 초기 설정값으로 IOU (Intersection over Union) threshold는 0.45, confidence score threshold는 0.50으로 하였다.

### 3.3 다중 카메라 네트워크 환경을 고려한 인원 합산

이 단계는 사람과 LOI의 이동관계를 인식하여 구역별 인원을 합산한다. 인원은 “서로 다른 영역을 구분하는 LOI를 지나는 상황”에 따라 합산되며, 앞서 설정한 LOI의 좌표값을 활용하여 추적한 사람 ID의 좌표가 LOI를 지나는지를 선분교차(Line-segment Intersection) 알고리즘으로 판정한다. 구체적으로 LOI의 두 개의 끝점( $p_1, p_2$ )과 ID의 직전 좌표( $p_3$ ), 그리고 현재 좌표( $p_4$ )가 주어진다고 가정하면, 다음과 같은 두 선분을 생각할 수 있다 (Eq. (1)).

$$\frac{p_1(x_1, y_1), p_2(x_2, y_2)}{p_1p_2, p_3p_4}, p_3(x_3, y_3), p_4(x_4, y_4) \quad (1)$$

이때 Cross product의 값을 나타내는 변수는 다음과 같이 정의할 수 있다(Eq. (2)).

$$\begin{aligned} d_1 &= \vec{p_1p_3} \times \vec{p_1p_4} \\ d_2 &= \vec{p_2p_3} \times \vec{p_2p_4} \\ d_3 &= \vec{p_3p_1} \times \vec{p_3p_2} \\ d_4 &= \vec{p_4p_1} \times \vec{p_4p_2} \end{aligned} \quad (2)$$

이처럼 한 선분의 양 끝점에서 다른 선분의 양 끝점을 종점으로 하는 두 벡터의 Cross product를 구하면, determinant의 부호에 따라 두 선분이 교차하는지를 파악할 수 있다(Fig. 4). 두 선분이 교차하는 경우  $d_1$ 과  $d_2$ 의 부호가 다르고,  $d_3$ 와  $d_4$ 의 부호가 다르게 된다. 따라서 다음과 같은 경우로 정리할 수 있다(Eq. (3)). 아래 값이 참인 경우, 두 선분은 교차한다고 판정한다.

$$((d_1 \times d_2) < 0) \text{ and } ((d_3 \times d_4) < 0) \quad (3)$$

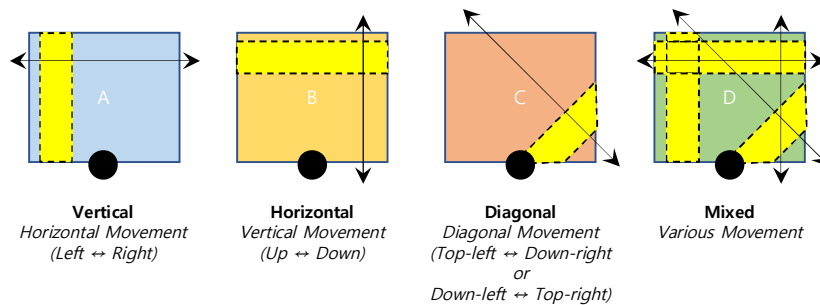


Fig. 3. LOI Drawing Cases

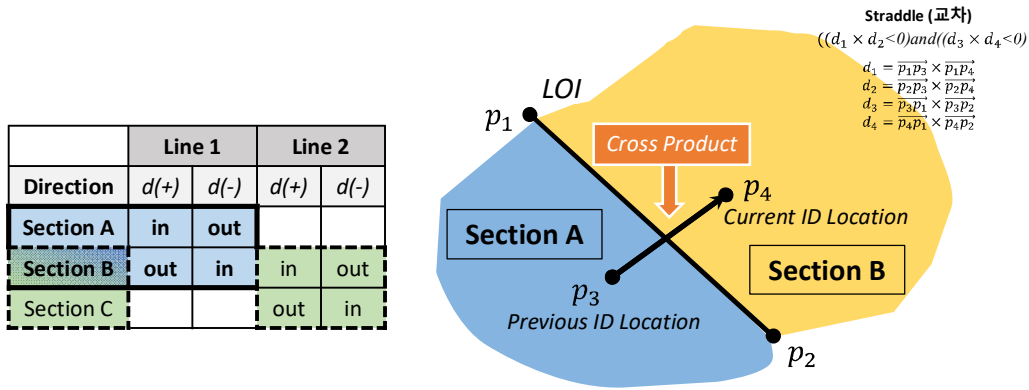


Fig. 4. Line-segment Intersection

일차함수를 사용하면 두 선분의 기울기가 매우 작은 경우 자료형으로 표현되는 과정에서 오류가 발생할 수 있지만 백터곱을 사용하면 이를 방지할 수 있다. 그리고 직전 좌표와 현재 좌표의 위치에 따라 진행 방향을 알 수 있기 때문에 구역 내 이동량을 쉽게 계산할 수 있다. 순차적인 공간적 위계(카메라 위계 코드)가 있는 LOI를 기준으로 생각해보면, LOI를 지나는 객체들이 집계된 결과에 따라서도 이동 정보를 파악할 수 있다(Fig. 4). 구역별 인원 합산은 LOI를 지나는 유무에 의해 판별한다. 외부에서 진입하는 경우 건물의 출입구 LOI를 안쪽으로 통과하는 경우만을 유효하게 인정했으며, 이후 구역 간 이동을 하지 않고 화면에 등장했다가 사라지는 경우는 해당 구역에 계속 머무는 것으로 계수했다.

#### 4. 실험 결과 및 분석

개발 모델을 검증하기 위해 서울대학교 건설환경공학부 5층 건물 내 CCTV 영상 데이터를 단위구역별로 수집하여 실험을 수행하였다. 개인의 이동정보를 활용하므로 서울대학교 생명윤리위원회(IRB)의 승인을 받아 인원 계수 실험을 진행하였다. 3개 시간대(오전, 정오, 오후)에 대해 각 15분 분량의 영상을 수집했으며 이를 집계하면 총 38대의 CCTV 카메라로부터 총 676,818장의 이미지를 수집하였다. 각 이미지의 해상도는 1280x1024이다. 그 후, 다크라벨

(DarkLabel) 소프트웨어(<https://darkpgmr.tistory.com/16>)를 이용하여 개별 이미지 프레임에 대해 객체의 유형, 위치를 기록하는 라벨링 작업을 수행하였고, 구축된 라벨링 데이터셋을 검증에 위한 정답값(Ground Truth)으로 활용하였다. 이 정답값에 기초하여 프레임별로 객체를 잘 인식했는지 평가하였고, LOI로 나눈 구역별 인원수의 기준값으로도 활용하였다. 실험은 Intel(R) Core(TM) i7-9700FK CPU @3.60GHz와 NVIDIA사의 GeForce RTX 2060, 32GB RAM이 설치된 프로그래밍 환경에서 수행되었다.

실험 결과, 개발된 모델은 영상에 나타난 사람을 정확히 탐지하고, 구역별 인원 수를 성공적으로 계수하였다. 구체적으로, 인식 결과를 정확도(accuracy)와 F1-Score로 표현하였는데, 각 시간대별로 평균 정확도는 오전 93.7%, 정오 88.4%, 오후 87.6%로 나타났고 평균 F1-score는 오전 81.9%, 정오 77.5%, 오후 78.2%로 나타났다(Table 1). 이는 화면의 객체 인식 정도를 사전에 작성된 라벨링 데이터셋의 정답값에 기초하여 산출되었다. 분석 속도를 나타내는 지표인 초당 분석 프레임 수(Frame per Second, FPS)는 평균 826으로 실시간 계수가 가능함을 보였다. 시간대별로 정확도가 다르게 나타난 이유는 카메라가 설치된 환경(각도, 조도, 기둥이나 벽과 같은 방해물 등)에 따라 인식 성능에 차이가 발생하는데, 재실자의 통행량이 늘어나고 줄어드는 장소가 시간에 따라 달라졌기 때문이다. 또한, 2명 이상의 인원이 서로 가까이 붙어서 이동하거

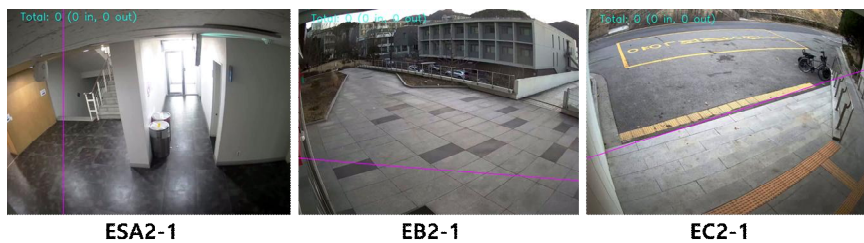


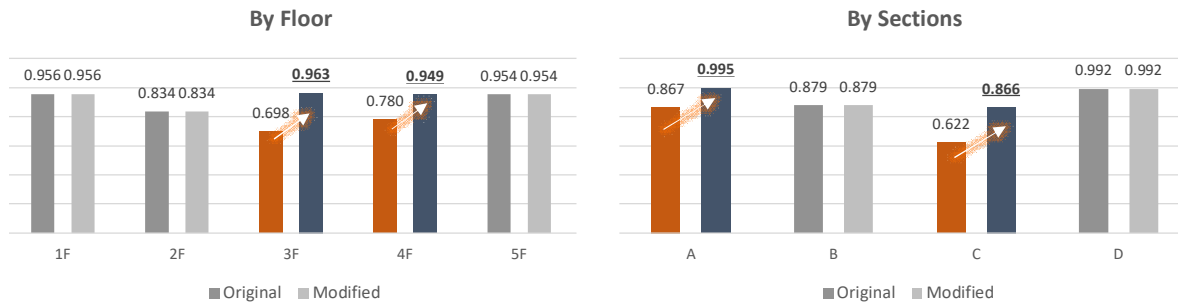
Fig. 5. Example of Camera FOV

**Table 1.** Counting Results by Floors and Sections

	Morning (09:00-09:15)	Noon (11:45-12:00)	Afternoon (18:00-18:15)
Total Frames	224,651	228,592	223,575
Ground Truth Frames	62,862	55,405	59,871
Predicted Frames	61,142	72,087	63,235
Average Accuracy	0.937	0.884	0.876
Average F1-score	0.819	0.775	0.782
Average FPS	10.14	9.13	5.50

**Table 2.** Counting Results by Floors and Sections

Category		Accuracy	MAE	RMSE
Floors	1	0.956	0.044	0.161
	2	0.834	0.275	0.584
	3	0.698	0.303	0.479
	4	0.780	0.221	0.335
	5	0.954	0.046	0.137
	<b>Average</b>	<b>0.844</b>	<b>0.178</b>	<b>0.339</b>
Sections	A	0.867	0.133	0.223
	B	0.879	0.179	0.423
	C	0.622	0.408	0.651
	D	0.992	0.008	0.072
	<b>Average</b>	<b>0.840</b>	<b>0.182</b>	<b>0.342</b>



**Fig. 6.** Example of LOI Adjustment Results

나 보행속도가 너무 빠른 경우 인식 성능이 떨어지기도 했는데, 이 경우 LOI의 위치를 재조정하여 화면에서 인식되는 객체의 위치와 속도를 고려함으로써 오차를 개선할 수 있었다. 구역별 계수 결과를 보편적으로 사용되는 평균 절대 오차(Mean Absolute Error, MAE)와 제곱근 평균 제곱 오차(Root Mean Square Error, RMSE)로 해석하였으며, 층별 평균 MAE와 RMSE는 각각 0.178, 0.339, 구역별 평균 MAE와 RMSE는 0.182와 0.342로 나타났다. 두 성능평가 지표는 실제 구역에 머무는 인원의 정답값과 합산한

예측값의 차이를 비교하여 도출되는데, 값이 작을수록 모델의 정답 값이 근사치로 맞춰지고 있음을 의미한다. 구역별 분류에 비해 층별로 분류한 결과의 성능이 좋았던 이유는 특정 구역에서 계수 오차가 크게 발생했기 때문으로 분석된다. 결과를 종합하면 Table 2와 같다. 추가로 LOI의 위치를 보정하는 parameter 조정을 통해 최적화를 실시한 결과 층별 평균 정확도는 84.4%에서 93.1%로, 구역별 평균 정확도는 84.0%에서 93.3%로 크게 상승하였다(Fig. 6). 이처럼 LOI 설정에 따라서도 모델의 객체 인식 정확도가 크게

향상될 수 있음을 확인했다.






## 5. 결론

본 연구는 컴퓨터 비전 알고리즘을 통해 건물 내 재실자 수를 계수하는 방법론을 제안하였다. 건물에 이미 설치된 CCTV 다중 카메라 환경에서 YOLOv4와 DeepSORT 알고리즘을 적용하여 다수의 사람을 감지 및 추적하고, 구역별 인원을 효과적으로 계수하기 위해 LOI를 활용하는 방법론을 제시하였다. 제안된 방법론은 5층 건물에서 수행된 실험을 통해 검증되었다. 제안된 방법론을 활용하여 평상시 건물의 층, 구역별 실시간 재실자 정보를 확보해둔다면 테러, 재난과 같은 특수한 상황이 발생했을 때 구조대원들에게 그 정보를 전달하여 신속한 구조 및 자원투입에 대한 의사결정을 지원할 수 있을 것이다. 향후 연구로 범비는 구역과 한산한 구역에 대한 가중치를 별도로 구분하여 모델의 정확도를 판별해볼 수 있다. 또한 시간대별, 일별, 월별, 주별, 요일별 등 다양한 시나리오로 구분하여 개발한 방법론을 적용하면 시설물 사용 현황 패턴을 분석해볼 수 있다. 나아가 도로, 터널, 교량 등의 시설물을 대상으로 해당 방법론을 적용한다면 시설물 내부에 머무르고 있는 인원을 파악하여 재난상황 발생 시 빠른 의사결정을 지원하는데 활용될 수 있을 것이다.

## 감사의 글

이 논문은 2021년도 서울대학교 융복합 연구과제와 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다(No.RS-2023-20241758).

## ORCID

Hwkyung Chun  <https://orcid.org/0000-0001-9512-058X>  
 Chanhyuk Park  <https://orcid.org/0009-0003-5201-6661>  
 Seokho Chi  <https://orcid.org/0000-0002-0409-5268>  
 Myungil Roh  <https://orcid.org/0000-0001-7972-6848>  
 Connie Susilawati  <https://orcid.org/0000-0001-5930-8331>

## References

Bochkovskiy, A., Wang, C. Y. and Liao, H. Y. M. (2020). "YOLOv4: Optimal speed and accuracy of object detection." *arXiv preprint*, <https://doi.org/10.48550/arXiv.2004.10934>.  
 Chae, S. U., Kwon, H. S., Park, S. R., Cho, W. H., Kwon, O. S. and Lee, J. S. (2020). "CCTV high-speed analysis algorithm for

real-time monitoring of building access." *Journal of the Korean Society of Hazard Mitigation*, KOSHAM, Vol. 20, No. 2, pp. 113-118, <https://doi.org/10.9798/KOSHAM.2020.20.2.113> (in Korean).  
 Denman, S., Fookes, C., Ryan, D. and Sridharan, S. (2015). "Large scale monitoring of crowds and building utilisation: A new database and distributed approach." *Proceedings of 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, Karlsruhe, Germany, pp. 1-6, <https://doi.org/10.1109/AVSS.2015.7301796>.  
 Dollar, P., Wojek, C., Schiele, B. and Perona, P. (2011). "Pedestrian detection: An evaluation of the state of the art." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, Vol. 34, No. 4, pp. 743-761, <https://doi.org/10.1109/TPAMI.2011.155>.  
 Javed, O., Shafique, K., Rasheed, Z. and Shah, M. (2008). "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views." *Computer Vision and Image Understanding*, Elsevier, Vol. 109, No. 2, pp. 146-162, <https://doi.org/10.1016/j.cviu.2007.01.003>.  
 Jung, S. P., Lee, H. Y. and Kim, J. W. (2018). "A study on the development of occupant density and walking pattern measurement techniques for emergency evacuation and safety in the railroad station: Focusing on information about pedestrians' use behaviors." *Journal of the Korean Society of Hazard Mitigation*, KOSHAM, Vol. 18, No. 1, pp. 125-135, <https://doi.org/10.9798/KOSHAM.2018.18.1.125> (in Korean).  
 Kalake, L., Wan, W. and Hou, L. (2021). "Analysis based on recent deep learning approaches applied in real-time multi-object tracking: a review." *IEEE Access*, IEEE, Vol. 9, pp. 32650-32671, <https://doi.org/10.1109/ACCESS.2021.3060821>.  
 Liu, A. S., Hsu, T. W., Hsiao, P. H., Liu, Y. C. and Fu, L. C. (2016). "The manhunt network: People tracking in hybrid-overlapping under the vertical top-view depth camera networks." *Proceedings of 2016 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, IEEE, Taipei, Taiwan, pp. 1-6, <https://doi.org/10.1109/ARIS.2016.7886632>.  
 Maddalena, L., Petrosino, A. and Russo, F. (2014). "People counting by learning their appearance in a multi-view camera environment." *Pattern Recognition Letters*, Elsevier, Vol. 36, pp. 125-134, <https://doi.org/10.1016/j.patrec.2013.10.006>.  
 Melfi, R., Rosenblum, B., Nordman, B. and Christensen, K. (2011). "Measuring building occupancy using existing network infrastructure." *Proceedings of 2011 International Green Computing Conference and Workshops*, IEEE, Orlando, FL, USA, pp. 1-8, <https://doi.org/10.1109/IGCC.2011.6008560>.  
 Park, C. and Chi, S. (2020). "Developing a zone-level people counting methodology using surveillance cameras for search and rescue efforts during building disasters." *Journal of the Spring Annual Conference of AIK*, AIK, Vol. 40, No. 1, pp. 421-424 (in Korean).  
 Perng, J. W., Wang, T. Y., Hsu, Y. W. and Wu, B. F. (2016). "The design and implementation of a vision-based people counting

- system in buses.” In *2016 International Conference on System Science and Engineering (ICSSE)*, IEEE, Puli, Taiwan, pp. 1-3, <https://doi.org/10.1109/ICSSE.2016.7551620>.
- Sharma, D., Bhondekar, A. P., Shukla, A. K. and Ghanshyam, C. (2018). “A review on technological advancements in crowd management.” *Journal of Ambient Intelligence and Humanized Computing*, Springer, Vol. 9, No. 3, pp. 485-495, <https://doi.org/10.1007/s12652-016-0432-x>.
- Sun, K., Zhao, Q. and Zou, J. (2020). “A review of building occupancy measurement systems.” *Energy and Buildings*, Elsevier, Vol. 216, 109965, <https://doi.org/10.1016/j.enbuild.2020.109965>.
- Wang, X. (2013). “Intelligent multi-camera video surveillance: A review.” *Pattern Recognition Letters*, Elsevier, Vol. 34, No. 1, pp. 3-19, <https://doi.org/10.1016/j.patrec.2012.07.005>.
- Wen, L., Lei, Z., Chang, M. C., Qi, H. and Lyu, S. (2017). “Multi-camera multi-target tracking with space-time-view hyper-graph.” *International Journal of Computer Vision*, Springer, Vol. 122, No. 2, pp. 313-333, <https://doi.org/10.1007/s11263-016-0943-0>.
- Wojke, N., Bewley, A. and Paulus, D. (2017). “Simple online and realtime tracking with a deep association metric.” *Proceedings of 2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, Beijing, China, pp. 3645-3649, <https://doi.org/10.1109/ICIP.2017.8296962>.
- Yu, S. I., Yang, Y. and Hauptmann, A. (2013). “Harry potter's marauder's map: Localizing and tracking multiple persons-of-interest by nonnegative discretization.” *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, OR, USA, pp. 3714-3720, <https://doi.org/10.1109/CVPR.2013.476>.
- Zhang, S., Zhu, Y. and Roy-Chowdhury, A. (2015). “Tracking multiple interacting targets in a camera network.” *Computer Vision and Image Understanding*, Elsevier, Vol. 134, pp. 64-73, <https://doi.org/10.1016/j.cviu.2015.01.002>.