보행자 깊이 정보를 이용한 군중 밀집도 추정

노유진[1], 이상민[2]
[1]광운대학교 인공지능융합학과 석사과정
[2]광운대학교 소프트웨어융합대학 교수
rohyujin55@gmail.com, smlee5679@gmail.com

The Crowd Density Estimation Using Pedestrian Depth Information

Yu-Jin Roh[1], Sang-Min Lee[2]
[1]Dept. of Artificial Intelligence Convergence, Kwangwoon University
[2]College of Software and Convergence, Kwangwoon University

요 약

다중밀집 사고를 사전에 방지하기 위해 군중 밀집도를 정확하게 파악하는 것은 중요하다. 기존 방법 중 일부는 군중 계수를 기반으로 군중 밀집도를 추정하거나 원근 왜곡이 있는 데이터를 그대로 학습한다. 이 방식은 물체의 거리에 따라 크기가 달라지는 원근 왜곡에 큰 영향을 받는다. 본 연구는 보행자 깊이 정보를 이용한 군중 밀집도 알고리즘을 제안한다. 보행자의 깊이 정보를 계산하기 위해 편차가 적은 머리 크기를 이용한다. 머리를 탐지하기 위해 OC-Sort를 학습모델로 사용한다. 탐지된 머리의 경계박스 좌표, 실제 머리 크기, 카메라 파라미터 등을 이용하여 보행자의 깊이 정보를 추정한다. 이후 깊이 정보를 기반으로 밀도 맵을 추정한다. 제안 알고리즘은 혼잡한 환경에서 객체의 위치와 밀집도를 정확하게 분석하여 군중밀집 사고를 사전에 방지하는 지능형 CCTV시스템의 기반 기술로 활용될 수 있으며, 더불어 보안 및 교통 관리 시스템의 효율성을 향상하는 데 중요한 역할을 할 것으로 기대한다.

1. 서론

다중밀집 사고는 밀집된 지역에서 사람, 차량 등 유동성을 가진 객체 간에 발생하는 대규모 사고를 의미한다. 특히 보행자의 사고는 전 세계에서 매년 발생하고 있으며, 공연장, 축제, 스포츠 경기 등 다양한 장소와 상황에서 발생한다. 따라서 심각한 인명피해와 부상을 초래할 수 있다. 이런 사고를 해결하기 위해 다중밀집 사고를 사전에 방지하는 연구가 필요하다.

최근, 정보기술(Information Technology, IT)의 발달로 컴퓨터 비전(Computer Vision)과 인공지능(Artificial Intelligence, AI)을 사용하여 군중 밀집도를 정확하게 파악하는 연구가 활발하게 진행되고 있다. 기존 연구에서는 인공지능 알고리즘이 군중 계수와 군중 밀집도를 학습하여 군중 밀도를 추정하였다 하지만 이러한 방법은 혼잡한 장면에서 원근 왜곡으로 인해 객체 간의 실제 거리와 이미지에서의 상대적 크기가 변화하는 문제에 직면하고 있다. 따라서 원근 왜곡을 고려하지 않은 기존 방법은 결과적으로 인공지능이 군중 밀도를 정확하게 추정하는데 어렵다.

본 논문에서는 보행자 깊이 정보를 이용한 군중 밀집도 알고리즘을 제안한다. 제안하는 알고리즘은 다음과 같이 2단계로 구분할 수 있다. 먼저, 보행자의 깊이 정보를 추정하기 위해 편차가 적은 머리를 탐지한다. 이를 위해, YOLOX[1]와 다중 객체 추적(Multi-Object Tracking, MOT) 알고리즘 중 하나인 OC-Sort(Observation-Centric SORT)[2]를 결합한 형태를 학습모델로써 활용하였다. 이후 탐지된 머리의 좌표와 실제 머리 크기, 카메라의 내각 정도를 이용하여 보행자의 깊이 정보를 추정한다. 다음으로, 깊이 정보를 기반으로 밀도 맵(Density map)을 추정한다.

본 연구에서는 OC-Sort가 실제 활용 목적에 적합한지를 판단하기 위해 MOT Challenge에서 제공하는 Head Tracking 21(HT21) 데이터를 활용하여 검증하였다. 보행자 머리 탐지를 실험한 결과 검증 데이터 세트에 대해 AP(Average Precision) 0.234,

AP_{50} 0.496의 성능을 보였다. 보행자 머리 추적은 MOTA(Multi-Object Tracking Accuracy) 61.8%의 성능을 보였다.

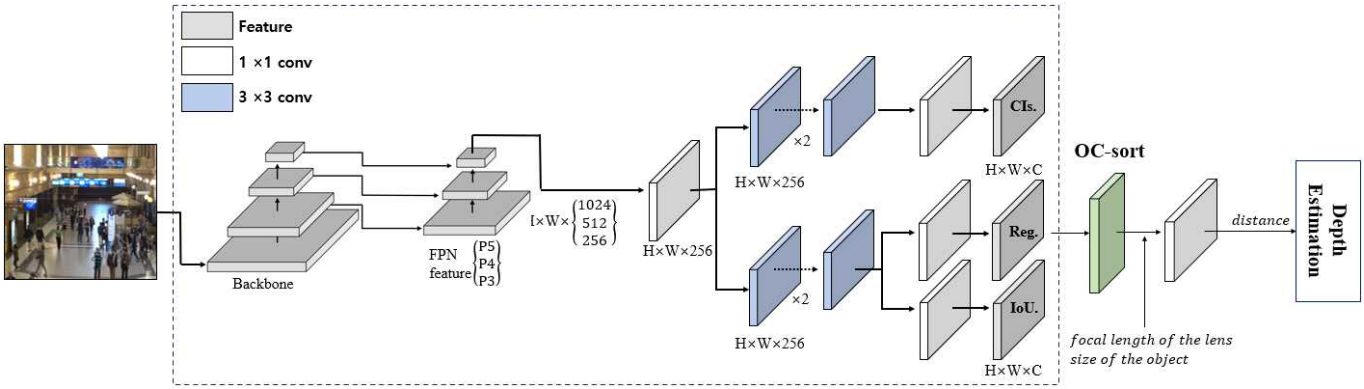


그림 1 제안하는 전체 알고리즘

2. 관련연구

2.1 다중객체추적(Multi-Object Tracking, MOT)

MOT은 연속적인 프레임에서 다중 객체를 동시에 탐지하고 인식하고자 하는 객체가 이동하는 경로상 이전 프레임(Frame)에서 탐지된 객체와 동일 객체인지를 인식하는 추적 기술을 지칭한다. 최근까지도 MOT을 수행하기 위한 많은 알고리즘이 개발되고 있다[3]. 그러나 대부분 임계값 보다 높은 경계박스와 경로조각(Tracklet)을 연결하여 동일 객체를 추적하게 된다. 이는 객체 간 중첩(Occlusion)되거나 자체 중첩(Self-occlusion)되는 경우 가려진 객체에 대한 경계박스(Bounding Box) 추정이 어려워지는 문제를 가진다. 중첩이 가장 적은 머리를 탐지하는 것이 보행자 자체를 탐지하는 것보다 좋은 성능을 보였다[4]. 본 연구에서는 CCTV의 데이터 특성을 고려하여, 중첩이 가장 적게 발생하는 머리의 경계박스를 추정하여 정확한 경계박스를 추정한다.

2.2 군중밀집도

기존 연구는 군중밀집도를 추정하기 위해서 군중계수를 기반으로 군중밀집도를 추정한다. 하지만 밀도가 높은 장면에서 군중 수를 추정하는 것은 균일하지 않은 규모 변화로 인해 매우 어려운 작업이다. 최근 군중계수의 취약점을 보완한 연구가 활발하다. 밀도 맵을 회귀하거나 전체 이미지에 객체 감지기를 적용하여 사람 수를 계산하는 대신 감지 및 회귀 모듈을 모두 사용하여 군중 수를 동시에 추정하는 DecideNet 알고리즘 연구가 있다[5]. 또한, 군중 수 분류와 밀도 지도 추정을 공동으로 학습하기 위한 새로운 종단 간 계단식 CNN(Convolutional Neural Network)에 대한 연구가 있다[6].

3. 제안 알고리즘

혼잡한 장면에서 군중 밀집도를 추정하는 것은 원근 왜곡으로 인해 매우 어려운 작업이다. 본 논문에서는 보행자 깊이 정보를 이용한 군중 밀집도 알고리즘을 제안한다. 제안하는 알고리즘은 2단계로 구성된다. 첫 번째 단계에서 CCTV 상에서 촬영된 보행자들의 머리를 학습하고 이후 카메라의 구조를 활용하여 보행자의 깊이를 추정한다. 두 번째 단계는 깊이 정보를 깊이 정보로 밀도 맵을 추정한다.

3.1 깊이 추정

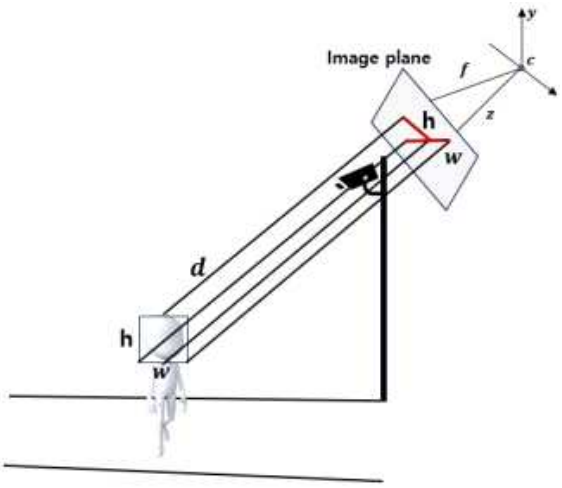
제안하는 알고리즘에 입력하기 위해 데이터를 전처리한다. 데이터 전처리를 위해 이미지의 밝기(Brightness)와 대비(Contrast), 선명화(Sharpening) 등의 처리를 거쳐 특징을 강조하였다. 또한, 알고리즘의 강건한 예측성능을 확보하기 위해 데이터증강(Data Augmentation) 기법을 적용하여 학습데이터를 구성하였다. 영상에 대한 데이터증강 시 좌우반전(Horizontal Flip), 이동(Translation) 등의 실제 데이터와의 유사한 특징을 나타낼 수 있는 증강 기법을 채택하였다. 또한, CCTV의 영상 특성을 반영하기 위해 일부 영상에 블러(Blur)와 크기 변경(Resizing)

을 추가로 적용하여 학습하였다.

다음으로 CCTV 카메라를 이용하여 보행자의 거리 정보를 추정한다. 컴퓨터비전 분야에서 영상에 대한 모든 기하학적 해석은 카메라 모델을 중점으로 한다. 카메라의 구조는 그림 2와 같다. 거리 정보를 구하기 위해서 카메라의 내각 정보뿐만 아니라 정확한 물체의 크기를 구하는 것이 필수적이다. 하지만 CCTV에 등장하는 수많은 보행자의 실제 크기를 구하는 것은 어려운 문제이다. 따라서 우리는 사람의 신체 중 크기의 편차가 가장 작은 머리를 이용한다. 성인 남성 20대를 기준으로 머리둘레에 대한 통계에 따르면, 머리둘레는 577.3mm로 편차가 ±0.6mm로 집계되었다. 우리는 통계를 기준으로 머리의 평균 크기를 실제 물체 크기로 정의한다. 카메라부터 머리까지의 거리 d를 계산하는 수식은 아래 식 1과 같다.

$$d = X \cdot \frac{f}{x} \quad (1)$$

d는 CCTV와 보행자 간의 거리, x는 이미지 크기, f는 렌즈의 초점거리 그리고 X는 머리둘레 길이의 평균값으로 정의한다. 이때 x는 OC-Sort의 출력값이다. 따라서 CCTV의 내각 정보인 f를 알면 구하고자 하는 카메라와 보행자 머리까지의 깊이 정보 d를 구할 수 있다. f는 렌즈 중심으로부터 초점거리까지의 거리를 의미하며, 규격에 따라 12.5mm로 정의한다.



(그림 2) 보행자 머리와 CCTV까지의 거리에 대한 CCTV 시스템 구조도

3.2 밀도 맵 추정

영상에서 객체는 원근 왜곡으로 인해 멀리 있는 객체가 가까이 있는 객체에 비해 작게 보이게 되고, 가까이 있는 객체는 크게 보이는 현상이 발생한다. 이는 실제 보행자 간의 거리의 왜곡을 발생시킨다. 따라서 해당 영역이 실제 밀집된 영역인지 원근 왜곡으로 인해 인구가 밀집되어 보이는지 판단이 어렵다. 이를 보정하기 위해 계산된 깊이 정보 값을 0과 1 사이의 값으로 정규화하고 밀도 맵을 생성한다. 밀도 맵에서 각 픽셀은 해당 위치의 보행자 밀도를 나타내는 값으로 구성되며, 밀도 맵에서 거리가 가까운 보행자는 높은 값으로 나타내고 거리가 먼 보행자는 낮은 값으로 나타내어 원근 왜곡에 따른 밀도를 보정한다.

4. 실험 결과

4.1 학습 환경 및 데이터 수집

본 연구에서 제안하는 방법론을 구현하기 위해 MOT 알고리즘인 OC-Sort에 보행자의 머리 위치를 학습시켰다. 우리는 본 연구를 위해 Intel i7-12700, RTX3080을 사용하였으며, 소프트웨어는 python 3.8.18, torch 1.13.1, OpenCV 4.8.0.76, numpy 1.23.2를 사용하였다.

OC-Sort 알고리즘을 학습시키기 위한 데이터셋으로 MOT Challenge에서 제공하는 Head Tracking 21(HT21)을 활용하였다[7]. HT21은 다양한 환경에서 촬영된 보행자들의 영상들로 구성되어 있으며, 보행자의 머리 경계박스를 탐지(Detection)하거나 추적(Tracking)하기 위한 정보들을 담고 있다. HT21은 학습 데이터셋과 테스트 데이터셋으로 구분되며, 학습 데이터셋은 4개의 동영상으로 총 5,741프레임의 1920x1080의 FHD(Full HD) 해상도를 가진다. 테스트 데이터셋은 5개의 동영상으로 총 5,723프레임의 FHD 해상도로 별도의 정답 데이터는 제공되지 않는다. 우리는 학습 데이터셋을 1:1 비율로 분할(Split)하여 학습 데이터셋과 검증 데이터셋으로 나누어 사용하였다.

4.2 성능평가 결과

4.1에서 학습시킨 OC-Sort를 검증 데이터셋으로 평가한 결과는 표 1, 2와 같다. 표 1은 보행자의 머리 경계박스를 탐지 성능이며, 표 2는 머리 경계박스 정보를 이용하여 추적하였을 때의 성능이다. 표 1을 살펴보면 큰 객체가 아닌 작은 객체(Tiny Object)를 찾다 보니 성능이 다소 낮아 보이나 AP_{50}의 성능이 0.496으로 유의미한 지표를 나타내는 것을 확인할 수 있다. 표 2에서는 MOTA가 61.8%를 보였다. HT21-03과 HT21-04 데이터셋에서 성능이 낮았는데

이는 해당 데이터셋에서의 경계박스 크기가 다른 데이터셋에 비해 작은 케이스가 많았기 때문으로 판단된다. 그림 3은 검증 데이터셋의 OC-Sort 결과를 시각화한 것이다.

<표 1> 보행자 머리 탐지 성능 결과

AP	AP_{50}	AP_{75}	AP_{small}	AP_{medium}
0.234	0.496	0.174	0.194	0.312

<표 2> 보행자 머리 추적 성능 결과

Dataset	MOTA↑	IDF1↑	IDs↓	MT↑	ML↑
HT21-01	68.8%	76.8%	27	39	10
HT21-02	63.1%	71.4%	356	322	70
HT21-03	59.2%	65.2%	460	178	85
HT21-04	59.6%	69.4%	277	101	57
Overall	61.8%	69.4%	1120	640	222

(그림3) 보행자 머리 탐지 및 추적 결과

5. 결론

본 논문은 군중 밀집도 향상을 위한 보행자 깊이 정보 추정을 제안한다. 이를 위해 다수의 보행자를 정확하게 탐지하기 위해 OC-Sort를 사용하여 머리를 학습한다. 핀홀 카메라의 구조를 활용하여 탐지된 머리와 카메라의 내각 정보로 카메라와 보행자까지의 깊이 정보를 추정한다.

본 연구에서는 OC-Sort가 실제 활용 목적에 적합한지 아닌지를 판단하기 위해 HT21 데이터를 활용하여 검증하였다. 실험 결과에서 보행자 머리 탐지의 평균 정밀도(AP)는 0.234로 나타났으며, 보행자 머리 추적에 대해서도 HT21 데이터셋의 다양한 시나리오에서 성능을 검증하였다. HT21의 검증 데이터셋에서의 전반적인 성능은 평균 61.8%로 나타났다. 제안 알고리즘은 혼잡한 환경에서 객체의 위치와 밀집도를 정확하게 분석하여 군중밀집 사고를 사전에 방지하는 지능형 CCTV시스템의 기반 기술로 활용될 수 있으며, 더불어 보안 및 교통 관리 시스템의 효율성을 향상하는 데 중요한 역할을 할 것으로 기대한다.

추후 연구에서는 제안하는 방법이 기존 방법에 비교하여 강건한지 평가하기 위해 원근 왜곡이 없는 보행자 밀도 데이터를 수집할 계획이다. 또한 머리 경계박스의 크기가 작기 때문에 tiny object detection 알고리즘을 사용하여 정확도를 높일 계획이다. 또한, 실시간 탐지가 중요하기 때문에 real-time object detection를 수행할 계획이다. 이를 고려하여 기존 Tracker에 RTMDet를 추가하여 추후 연구를 진행하고자 한다.

참고문헌

[1] GE, Zheng, et al. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430, 2021.

[2] Bewley, Alex, et al. "Simple online and realtime tracking." 2016 IEEE international conference on image processing (ICIP). IEEE, 2016.

[3] Lyu, Chengqi, et al. "Rtmdet: An empirical study of designing real-time object detectors." arXiv preprint arXiv:2212.07784 (2022).

[4] Sundararaman, Ramana, et al. "Tracking pedestrian heads in dense crowd." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[5] Liu, J., Gao, C., Meng, D., & Hauptmann, A. G. (2018). Decidenet: Counting varying density crowds through attention guided detection and density estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5197-5206).

[6] Sindagi, V. A., & Patel, V. M. (2017, August). Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1-6). IEEE.

[7] SHAO, Shuai, et al. Crowdhuman: A benchmark for detecting human in a crowd. arXiv preprint arXiv:1805.00123, 2018.