

IDEA: 실내 밀도추정 연구를 위한 ARMA3 기반 데이터 생성 프레임워크

김성현* 김예찬* 김시현 전문구

GIST 전기전자컴퓨터공학부

{ha307702, yechankim, sihyunkim2}@gm.gist.ac.kr, mgjeon@gist.ac.kr

IDEA: A Framework to Build Data for Indoor Density Estimation with ARMA3

Suncheon Kim*, Yechan Kim*, Sihyun Kim and Moongu Jeon

School of Electrical Engineering and Computer Science, GIST

요약

군중 밀도추정은 공공 안전을 위한 모니터링, 도시 인프라 기획 등에 유용하기에 최근 컴퓨터 비전 커뮤니티에서 활발하게 연구가 이루어지고 있다. 이러한 기술은 실내 환경에서도 유의미하게 활용될 수 있지만, 현재 대부분의 선행 연구에서는 실외 환경을 중심으로 구축된 데이터셋을 주로 다루고 있다. 이는 실내 환경에 특화된 밀도추정 데이터셋의 부재로 인한 것으로, 기존의 밀도추정 모델은 실내 환경에서의 추론 정확도가 떨어지는 문제점이 존재한다. 이에 따라 본 연구에서 우리는 실내 밀도추정을 위한 자동화된 가상 데이터 생성 프레임워크 IDEA (Indoor Density Estimation with ARMA3)를 제안한다. 제안하는 프레임워크의 특징점은 다음과 같다: 1) 완전히 자동화된 밀도추정 영상 및 어노테이션 쌍의 생성이 가능하다; 2) 사용자의 실내 환경 모델링 기능을 제공한다; 3) 혼합 확률분포 (Mixture Distribution) 기반의 다양성 높은 군중 밀도 모델링이 가능하다; 4) 다중 카메라 시점 (Multi-camera View)에서의 영상 취득이 가능하다. 제안하는 IDEA는 최소 비용으로 현실감과 다양성을 모두 확보한 가상 데이터 (Synthetic Data) 생성을 지원하므로, 향후 실내 밀도추정 연구 고도화에 크게 기여할 것으로 기대된다.

1. Introduction

군중 밀도추정은 최근 컴퓨터 비전 커뮤니티에서 큰 관심을 얻고 있는 분야이다. 이 기술은 인구 과밀집에 따른 사고를 예방하고, 자원시설을 전략적으로 배치하고, 더 나아가 전염성 질병의 확산 경로를 예측하는 것과 같이 다양한 차원에서 공공 안전 및 발전을 위해 활용될 수 있다. 이에 따라 최근 밀도추정과 관련한 다수의 선행 연구가 제안된 바 있으며, UCF_CC_50 [1], ShanghaiTech-Part A [2], UCF-QNRF [3], NWPU-Crowd [4], JHU-Crowd [5]와 같은 벤치마크 데이터셋에서 CNN 및 트랜스포머 기반의 신경망을 바탕으로 큰 성능 향상을 달성한 바 있다.

하지만 기존의 밀도추정 모델은 실내 환경에서 취약하다. 밀도추정 모델 훈련을 위한 종래의 벤치마크 데이터셋들은 야외 환경을 중심으로 수집·가공되어, 이러한 데이터로 훈련된 모델은 실내 조건에서의 일반화 성능이 떨어지기 때문이다. 한편, 실내 환경에 특화된 다중 카메라 시점의 데이터셋은 우리가 아는 한 존재하지 않는다. [6, 7]은 이를 극복하고자 컴퓨터 비전에서 나아가 WiFi 신호 기반의 군중 밀도추정 방법을 제안하였다. 하지만 이는 모바일 단말을 보유하고 있지 않거나, 다수의 단말을 보유한 사람들로 인해 예측의 한계가 존재한다. 이에 따라 본 고에서는 시각지능 관점에서 군중 밀도추정에 집중한다.

실내 밀도추정은 자연광 및 조명 유무에 따른 불균일한 조명 조건, 각종 가벽 등에 따른 사각 지대, 복도와 같은 통로에서의 유동 인력에 따른 과밀집 등의 난제 요인이 존재하며, 이를 위해 단일이 아닌 다중 카메라 시점의 영상 기반의 연구가 수행되어야 한다. 한편, 이러한 모든 경우를 포괄하는 대규모 딥러닝 모델 학습용 데이터를 실제 세계에서 확보하는 것은 초상권 및 어노테이션 비용 등의 문제로 용이하지 않다.

따라서 본 고에서 우리는 실내 밀도추정 연구를 견인하기 위한 가상 데이터 생성 프레임워크 IDEA (Indoor Density Estimation with

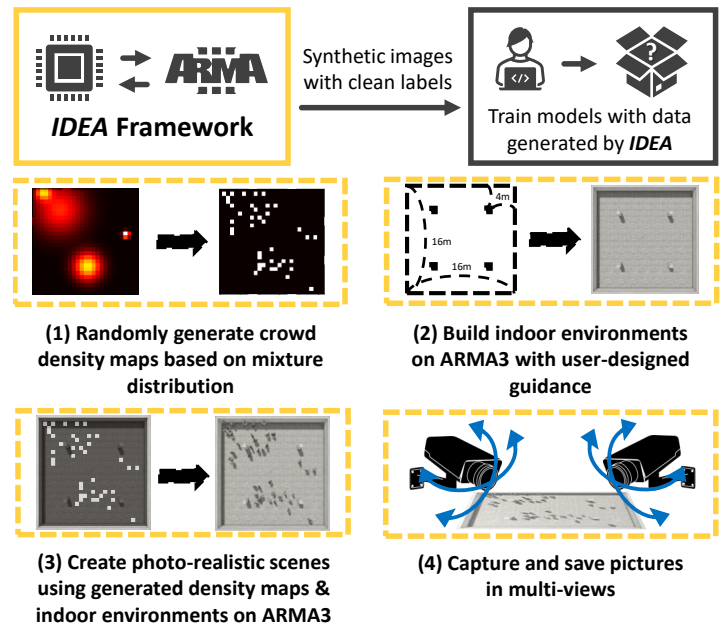


그림 1: 제안하는 실내 밀도추정용 데이터 생성을 위한 IDEA (Indoor Density Estimation with ARMA3) 프레임워크의 개관

ARMA3)를 제안한다. ARMA3는 Bohemia Interactive 사가 개발한 군사 전술 시뮬레이션 게임이다. ARMA3는 정교한 물리 시뮬레이션을 지원하는 Real Virtuality 게임 엔진을 바탕으로 고품질의 사용자 맞춤형 시나리오 및 콘텐츠 생성을 가능케한다. 우리는 ARMA3를 바탕으로 다음과 같은 이점을 갖춘 새로운 프레임워크를 개발하여 공개한다:

1. 컴퓨터 시뮬레이션 기반의 밀도추정 영상 수집 및 레이블링의 자동화 실현을 통한 데이터 어노테이션 비용 제거;
2. 사용자의 실내 환경 모델링 기능 지원;
3. 혼합 확률분포 기반의 다양성 높은 군중 밀도분포 생성;
4. 동일 장소에 대한 다중 카메라 시점의 영상 취득 기능 제공.

1) 이 성과는 2024년 정부(과학기술정보통신부) 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2023R1A2C2006264) (* denotes equal contribution.)

본 논문은 다음과 같이 전개한다. 2장은 제안하는 **IDEA** 프레임워크의 개관 및 주요 알고리즘을 설명하고 추가적 고려사항에 대한 심층 논의를 진행한다. 3장은 제안 방식으로 생성한 밀도추정 모델 지도학습용 데이터의 품질을 고찰한다. 4장은 제안 방법의 추가 활용 및 개선 방향에 대해 논의하며 논문을 끝낸다.

2. Proposed Approach

본 장에서는 실내 밀도추정 연구용 가상 데이터 생성 프레임워크 **IDEA**를 살펴본다. 우선, 그림 1에서 핵심에 해당하는 단계 (1), (2)를 중심으로 **IDEA**의 아이디어를 설명한다. 이후 실제 구현 관점에서 주목해야 할 이슈들을 논의한다.

2.1 Randomly Generate Crowd Density Maps based on Mixture Distribution

알고리즘 1: 혼합 확률분포 기반 임의의 군중밀도 맵 생성 코드

```
1 # Step 1: Define distribution function(s) to use
2 def dist_1() -> np.array:
3     # (Example) 2D Gaussian distribution
4     mean_x = np.random.randint(w)
5     mean_y = np.random.randint(h)
6     std = np.random.rand()
7     X, Y = np.meshgrid(np.arange(w), np.arange(h),
8                         indexing='ij')
9     out = np.exp(-((X-mean_x)**2+(Y-mean_y)**2) /
10                 (2*std**2))
11     return out
12 # Step 2: Call get_density_map()
13 def get_density_map() -> np.array:
14     dists = [dist_1, ...] # can be extended
15     dmap = np.zeros((w, h))
16     for _ in range(np.random.choice([1, ..., n])):
17         alpha = np.random.rand() * n
18         dmap += alpha * np.random.choice(dists)()
19     normalized = dmap / np.sum(dmap)
20     scaled = normalized * p
21     mask = np.random.rand(*dmap.shape)
22     one_hot = np.where(scaled >= mask, 1, 0)
23     return one_hot
```

위 알고리즘 1은 다양한 군중 밀도맵 (Crowd Density Map)을 임의로 생성하기 위한 파이프라인에 관한 것이다 (그림 1-(1) 참조). 편의상 우리는 Python3 및 Numpy 스타일로 의사코드를 제시한다. 우리는 혼합 확률분포 (Mixture Distribution)를 바탕으로 군중 밀도맵을 생성하고, 2.3 절에서 이를 바탕으로 가상의 군중 영상 (Crowded Scene)을 제작할 것이다. 혼합 분포는 여러 다른 확률 분포의 가중 조합을 바탕으로 복잡한 데이터 구조를 모델링하는데 효과적인 도구 [8]이기에, 우리의 방법은 이를 바탕으로 실세계의 군중 분포를 최대한 근사하여 가상 데이터의 현실성을 강화한다.

위 알고리즘을 살펴보기에 앞서, 표기 약속은 다음과 같다. w 와 h 는 사람의 평균 어깨 너비 (이하 sw 라 함)로 나눈 공간 바닥의 가로 및 세로 길이이다. 예로, sw 를 0.6m로 가정할 때, 16m의 공간 바닥 가로 길이에

대응하는 w 는 $\frac{16}{0.6} \approx 27$ 이다. n 은 분포의 모드 (Mode) 개수이다. p 는 영상에 출현할 사람 수이다. n 과 p 는 적절한 임의의 자연수로 설정한다.

우선, (Step 1)은 확률분포 함수를 정의하는 부이다. 혼합 확률분포 생성 과정에서 활용하기 위한 1개 이상의 함수를 정의하여야 한다. 본 고에서는 2차원의 정규 분포함수를 예시로 제시한다 (`dist_1()`). 반드시 2차원 평면 공간에서 확률분포를 정의해야 하며, 배치 단위 연산을 위해 행렬 (2차원 배열)로 계산 결과를 반환하도록 한다.

(Step 2)에서는 임의의 군중 밀도맵을 생성하여 반환해주는 함수 `get_density_map()`이 제시된다. 사용자는 Line 14에서 `dists` 리스트에 (Step 1)에서 정의한 1개 이상의 모든 함수명을 열거한 뒤, `get_density_map()` 함수를 호출하여 사용한다. Line 15은 모든 값이 0으로 초기화된 공간 바닥 넓이의 2차원 배열 `dmap`을 선언한다. 이후, Line 16-19에서 1 이상 n 이하의 임의의 개수의 분포의 가중 합을 통해 혼합 분포를 만들고, 이를 `dmap`에 저장한다. Line 19는 `dmap`의 합이 1이 되도록 `dmap`을 정규화 (Normalization)하고, Line 20에서 사용자가 설정한 사람 수 p 로 이를 재조정 (Rescaling)한다. Line 21-22는 0과 1 사이의 균등 분포를 따르는 난수 배열을 활용하여 밀도맵을 이진화 (Binary Thresholding)한다. 특히, Line 21의 의미는 밀도맵에 일종의 잡음 (Noise)을 추가하여 무작위성을 향상하기 위함으로 이해할 수 있다. 최종적으로 이진화된 밀도맵 (`one_hot`)은 1 또는 0의 값을 갖는데, 1에 해당하는 픽셀에만 사람을 배치하도록 한다.

2.2 Build Indoor Environments on ARMA3 with User-Designed Guidance

ARMA3에 존재하는 다양한 구조물들을 적절히 배치하여 가상 실내 환경의 구현이 가능하다. 예로, 그림 1의 (2)에서와 같이, 사용자는 우리의 프레임워크를 사용하여, 16m²의 공간을 콘크리트 벽으로 생성한 뒤, 이웃한 두 벽으로부터 4m 떨어진 지점마다 기둥을 설치할 수 있다. ARMA3는 다양한 재질의 벽, 기둥, 장애물들을 원하는 위치와 방향으로 배치할 수 있다. 본 고에서는 다루지 않지만, ARMA3에서 더욱 현실 세계와 같은 실내 환경을 구현하고 싶다면, 실세계 영상을 물체 표면에 삽입하는 방식으로 실내 환경의 모델링이 가능함을 알린다.

2.3 Pivotal Issues Requiring Deliberation

본 절은 **IDEA** 프레임워크의 구현·배포 관점에서 발생할 수 있는 추가적인 고려 사항에 대해 심층 논의한다².

- (Q1) 3차원 가상 환경의 임의 좌표 (x, y, z)를 2차원 이미지 평면상의 좌표로 변환하려면?
 - (A1) 캘리브레이션을 통해 카메라 파라미터를 계산하고, 이를 통해 좌표 변환을 수행하는 것이 원칙이다. 다만, ARMA3의 경우, 내장된 게임 엔진에서 이를 위한 기능을 지원하므로 별도의 자체 구현을 요구하지 않는다.
- (Q2) 겹침 등의 이유로 일부 사람이 가려진 (Occlusion) 경우, 어노테이션은 어떻게 수행하나?
 - (A2) 기본적으로 군중 밀도추정 태스크에서는 사람의 머리 (Head)의 중앙점을 레이블로 활용한다. 따라서, 특정 인물

2) 본 고는 페이지 분량 제약으로 프로그래밍 구현 수준에서의 구체적이고 미시적인 논의 대신 거시적인 논의를 실시한다. 추후 후속 논문에서 더욱 자세한 프로세스를 설명할 예정이다.



그림 2: 제안하는 **IDEA** 프레임워크로 생성한 실내 밀도추정 데이터 샘플 (단, ●은 레이블)

의 머리가 여러 이유로 가려진 경우에는 이를 식별하는 코드를 추가적으로 구현하여, 해당 레이블을 출력하지 않게 한다.

- (Q3) 2.1절에서 생성한 밀도 맵을 바탕으로 ARMA3 환경에서 사람을 배치한다고 하자. 만약, 2.1절 밀도 맵의 특정 픽셀이 ARMA3 환경에서 장애물 (가벽 등)에 해당할 경우는?
 - (A3) ARMA3의 경우, 장애물과 객체 간의 교차 (Intersection)를 자동으로 방지해주는 기능이 내장되어 있다. 이러한 교차 상황에서, 사람의 배치 위치는 자동으로 적절하게 수정되고, 수정된 위치를 다시 변환받아서 객체가 실제 배치된 위치 정보를 ARMA3에서 획득할 수 있다.
- (Q4) 다중 카메라 기반의 다중 시점 영상은 어떻게 취득하나?
 - (A4) 본 프레임워크에서는 게임 상의 시간을 정지한 뒤, 사용자 입력을 통해 임의의 개수의 다중 위치에 카메라를 설치한 뒤, 영상 및 이에 대응하는 어노테이션 쌍을 취득하도록 지원한다. 예로, 사용자는 구면 좌표계를 활용하여 다중 CCTV 환경을 구현하고, 이를 바탕으로 데이터를 취득할 수 있다.

3. Qualitative Analysis

그림 2는 우리가 개발한 **IDEA** 프레임워크로 생성한 실내 밀도추정 연구를 위한 데이터 샘플이다. 영상 위에 ●은 배치된 사람 객체의 Head point이며, 밀도추정 모델 훈련을 위한 레이블로 사용된다. 영상 촬영시 카메라 각도는 구면 좌표계의 θ 와 ϕ 를 이용하여 설정할 수 있는데, 그림 2는 θ 가 23.5°와 43.5°, ϕ 가 30°, 45°, 60°일 때 조건으로 촬영된 결과물이다. 위 그림 2에서 알 수 있듯이, 우리의 프레임워크를 이용하면 자유로운 카메라 시점에서 현실 세계 CCTV와 유사한 조건에서 데이터를 취득할 수 있다. 따라서, 제안하는 **IDEA**는 향후 실내 밀도추정 연구를 효과적으로 견인할 것으로 기대된다.

다만, ARMA3는 군사 게임이므로, 군복을 착용한 남성 유닛이 대부분이며 모든 유닛의 체형이 동일하다. 본 연구의 데이터 생성 과정에서는 해당 문제를 한정된 민간인 유닛 사용과 유닛 스케일의 변동성을 주

는 식으로 보완하였으나 현실 데이터에 비해 데이터의 다양성을 저해할 수 있는 요인이 된다. 이런 문제를 해결하기 위해 우리는 추후 온라인 커뮤니티에 무료 공개된 애셋을 함께 활용할 것이다.

4. Conclusion

본 논문은 ARMA3를 이용하여 실내 군중 밀도추정 연구를 위한 가상 데이터 생성 프레임워크 **IDEA**를 제안하였다. 또한, 이를 바탕으로 구축한 데이터를 제시하고 그 품질에 대한 정성적 평가를 실시하였다. 이는 학술적으로 공개된 최초의 연구 시도이며, 유연한 확장성 및 범용성을 갖춘 방법임을 강조한다. 후속 연구에서는 **IDEA**를 바탕으로 데이터셋을 구축하고, 이를 바탕으로 한 모델 훈련 기법을 개발할 것이다.

참고 문헌

- [1] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multi-scale counting in extremely dense crowd images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2547–2554, 2013.
- [2] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 589–597, 2016.
- [3] H. Idrees, M. Tayyab, K. Athrey, D. Zhang, S. Al-Maadeed, N. Rajpoot, and M. Shah, "Composition loss for counting, density map estimation and localization in dense crowds," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 532–546, 2018.
- [4] Q. Wang, J. Gao, W. Lin, and X. Li, "Nwpu-crowd: A large-scale benchmark for crowd counting and localization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 6, pp. 2141–2149, 2020.
- [5] V. A. Sindagi, R. Yasarla, and V. M. Patel, "Jhu-crowd++: Large-scale crowd counting dataset and a benchmark method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 5, pp. 2594–2609, 2020.
- [6] Y. Yuan, "Crowd monitoring using mobile phones," in *2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 1, pp. 261–264, IEEE, 2014.
- [7] X. Tang, B. Xiao, and K. Li, "Indoor crowd density estimation through mobile smartphone wi-fi probes," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 50, no. 7, pp. 2638–2649, 2018.
- [8] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.