Edge-Based Multimodal Crowd Monitoring System for Outdoor Environments

Taemin Hwang*
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Seongnam, Republic of Korea
taemin.hwang@keti.re.kr

Won Gi Choi
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Seongnam, Republic of Korea
cwk1412@keti.re.kr

Sohyeon Kim
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Seongnam, Republic of Korea
shk980903@keti.re.kr

Jinyoung Lee
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Seongnam, Republic of Korea
jylee@keti.re.kr

Young Hwan Jeong
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Seongnam, Republic of Korea
cjstntjd@keti.re.kr

Minjoon Kim
School of Electrical and Computer Engineering
University of Seoul
Seoul, Republic of Korea
mjoon@uos.ac.kr

***Abstract*— This study proposes a portable, edge-based multimodal system for real-time crowd monitoring in outdoor performance. The device integrates vision-based headcounting using a night-vision camera with wireless signal analysis via BLE advertisements and WiFi probe requests. Built on a Raspberry Pi 5 with LTE module and battery power, the system is designed for flexible deployment in temporary, infrastructure-limited environments. During field trials, data were collected at regular intervals. The results show that BLE MAC address counts and camera-based headcounts generally increased during crowd entry, with BLE trends closely tracking temporal changes. Also, wireless signal analysis complemented visual sensing under low-light conditions. These findings suggest that combining visual and wireless modalities enhances robustness in dynamic monitoring scenarios.**

Keywords*— *Multimodal Crowd Monitoring, Edge-Based Sensing System

I. INTRODUCTION

Crowds commonly gather in public places such as airports, stadiums, parks, and transportation hubs, requiring robust systems to ensure public safety during large-scale events [1,2]. Outdoor performance venues, such as concerts and festivals, pose unique safety challenges due to their temporary setups, irregular spatial layouts, and lack of permanent monitoring infrastructure. These conditions complicate the deployment of conventional systems like CCTV or fixed IoT sensors, and often leave organizers reliant on manual monitoring, which is insufficient for real-time response.

Recent research has explored sensor-based crowd monitoring using vision and wireless signal modalities. Vision-based approaches utilizing deep learning models such as YOLO offer high spatial resolution and can estimate headcounts effectively. However, they are sensitive to occlusion, viewing angles, and low-light conditions. In contrast, wireless signal-based methods—particularly those using Bluetooth Low Energy (BLE) and WiFi probe requests—offer privacy-preserving alternatives that are more resilient to environmental variability. By analyzing packets broadcasted from smartphones or wearable tags, these methods can infer the number of people in an area [3-5].

This study introduces an edge-based, real-time crowd safety monitoring system that fuses these complementary sensing modalities. The system comprises a compact device equipped with a Raspberry Pi 5, a night-vision camera, BLE/WiFi receivers, and an LTE communication module powered by a rechargeable battery. This portable design enables flexible, autonomous deployment without requiring external power or network infrastructure, making it ideal for temporary outdoor venues. Alerts can be issued through LTE to notify safety personnel in real time.

II. PROPOSED SYSTEM

The proposed system consists of edge devices for multimodal crowd monitoring, as shown in Fig. 1. It features a multimodal sensing platform that combines a camera module for vision-based head counting and a BLE/WiFi receiver for wireless signal analysis. The system employs a Raspberry Pi 5 as the core processing unit, performing real-time YOLO-based inference and signal processing. It is equipped with an LTE modem for remote communication and a rechargeable battery to support autonomous operation in outdoor environments. Its modular and design allows for quick field deployment without the need for external infrastructure, making it especially suitable for temporary performance settings.

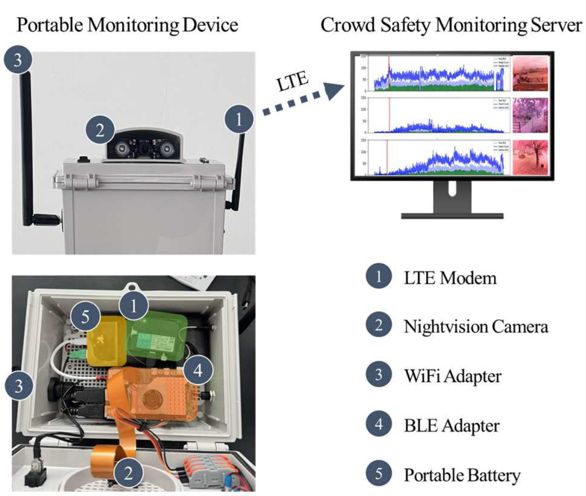

Fig. 1. Edge-based multimodal crowd monitoring system

A. Camera-Based Crowd Counting

The system uses a night vision Picam module to capture video frames at regular intervals. A YOLO-based deep learning model performs real-time object detection on edge devices to count the number of visible heads in each frame. The count is timestamped and stored locally, enabling temporal analysis of crowd changes.

B. Wireless Signal Analysis

To complement camera-based analysis, we analyze the wireless signals from nearby devices for estimating crowd presence more robustly.

- **BLE**: The BLE receiver scans for advertising packets broadcasted by nearby beacons or smartphones using the *bleak* python package. The number of unique MAC addresses detected over time serves as a proxy for crowd density [4,5].
- **WiFi***:* Probe request packets emitted by mobile devices are captured using *tcpdump*. These packets are parsed to extract sender MAC addresses, and the total count per unit time is used to estimate user presence [3].

C. Crowd Analysis

Camera-based and wireless data are synchronized and fused based on their timestamps. A weighted combination of headcount and signal-based measurements is used to estimate overall crowd density. This multimodal approach enhances robustness against occlusion or wireless blind spots, ensuring accurate and timely detection of overcrowding.

TABLE I. DATA STRUCTURE FOR CROWD ANALYSIS

Field	Description	Type
timestamp	Data recording time	string
wifi mac count	Number of WiFi MAC addresses	integer
ble mac count	Number of BLE MAC addresses	integer
head count	Number of detected objects	integer

III. EXPERIMENTAL RESULTS

To validate the proposed system, we conducted a field deployment at an outdoor performance venue during a live event. The multimodal device was positioned at a fixed location to monitor the audience flow over time. Data was collected with the following intervals: BLE signals were captured every second, camera frames were recorded every 5 seconds, and Wi-Fi packets were analyzed every 5 minutes.

As illustrated in Figure 2, during the entry phase of the event, we observed a significant increase in both visual headcount and the number of BLE MAC addresses detected. This correlation confirmed that BLE density can serve as a reliable proxy for crowd size. Furthermore, when a moving average filter was applied to the BLE MAC count, its trend closely matched the variation in headcount over time, indicating the effectiveness of signal smoothing for temporal analysis. Wi-Fi packet trends showed a noticeable lag due to the five-minute aggregation interval, making them appear temporally shifted compared to BLE and camera data. However, they still captured the general trend of increasing and decreasing crowd density.

Notably, during the exit phase in the evening, visual detection became unreliable due to low-light conditions. Despite the camera's limitations, both BLE and Wi-Fi signals remained elevated, suggesting the continued presence of a dense crowd. This result highlights the complementary nature of wireless signal analysis in scenarios where vision-based methods are constrained. Overall, the experimental findings support the feasibility and robustness of the proposed multimodal system for real-time crowd monitoring in outdoor, infrastructure-limited environments.

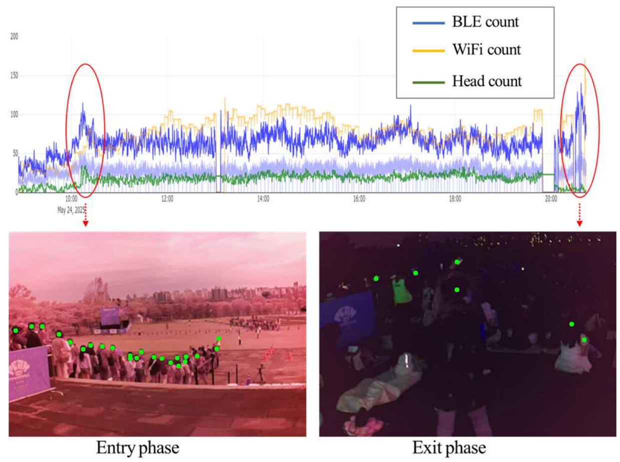

Fig. 2. Experimental results conducted in an outdoor performance

IV. CONCLUSION

This study presented a portable, edge-based system that fuses camera-based headcounting with BLE and Wi-Fi signal analysis for crowd monitoring. Field experiments showed that BLE trends closely align with visual data, and that wireless signals remain effective when low-light conditions hinder vision. The system enable rapid deployment without fixed infrastructure. These results highlight the value of multimodal sensing for robust, real-time crowd safety management. Future research will focus on developing machine learning models to predict crowd surges based on historical signal patterns to further improve early warning accuracy.

ACKNOWLEDGMENT

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency(KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2024 (Project Name: Development of AI-based performance site crowd concentration accident prediction and real-time response platform technology, Project number: RS-2024-00339321)

REFERENCES

[1] S. Lamba and N. Nain, ‘‘Crowd monitoring and classification: A survey,’’ in *Advances in Computer and Computational Sciences (Advances in Intelligent Systems and Computing)*. 2017, doi: 10.1007/978-981-10-3770-2_3.

[2] Still. K., "Crowd Safety and Crowd Risk Analysis," *Static Crowd Density Visuals*, 2019. (https://www.gkstill.com/Support/crowd-density/100sm/index.html)

[3] K. Gebru et al., "Edge-based passive crowd monitoring through WiFi Beacons," *Computer Communications*, vol. 192, pp. 163–170, 2022. doi: 10.1016/j.comcom.2022.06.003

[4] A. Basalamah, "Sensing the Crowds Using Bluetooth Low Energy Tags," *IEEE Access*, vol. 4, pp. 4225–4234, 2016. doi: 10.1109/ACCESS.2016.2594210

[5] Y. Qamaz et al., "Experimental evaluation of using BLE beacon for outdoor positioning in GPS-denied environment," *AGILE GIScience Series*, vol. 3, p. 13, 2022. doi: 10.5194/agile-giss-3-13-2022