HajjFlow: An Edge-Based Computer Vision System for Crowd Monitoring During Hajj

Abdulrahman Hameed Alosaimi
Alhassan Mohammed Asiri
Department of Computer Science
King Khalid University

***Abstract*—Managing crowds during the Hajj pilgrimage is a major safety and operational challenge due to the presence of millions of pilgrims in limited spaces. Although many CCTV cameras are deployed, most crowd monitoring is still performed manually, which limits scalability and timely response. This paper presents *HajjFlow*, an edge-based computer vision system for real-time crowd detection and density estimation. The system uses YOLO for person detection in low-to-medium density scenes and CSRNet for density estimation in highly crowded scenes. Video processing is performed at edge devices, while only numerical metrics are transmitted to a central dashboard for visualization. The proposed system aligns with Saudi Vision 2030 and the National Strategy for Data and Artificial Intelligence (NSDAI) by demonstrating the use of artificial intelligence to improve public safety and decision support during large-scale events.**

I. Introduction

One of the largest recurring mass gatherings in the world is the annual Hajj pilgrimage, attracting millions of pilgrims each year. The ability to manage very densely populated areas of pilgrims in an orderly and safe manner is essential for the operation of safe and successful Hajj, including safety before, during and after the Hajj, efficient operation and timely actions regarding congestion and delays.

There is currently a reliance on human operators manually monitoring the video feeds of approximately 5,000 CCTV cameras to assist in crowd monitoring, including for emergency situations. However, there are limitations associated with this methodology based upon the individual limitations of each operator, including the fatigue of operators, their reaction times and their ability to monitor large-scale crowd dynamics.

The Saudi Vision 2030 and the National Strategy for Data and Artificial Intelligence (NSDAI) promote the implementation of digitisation and advanced digital technologies in public services and to improve safety [4], [5]. In this setting, HajjFlow is being developed to be an edge-based computer vision system that democratises crowd monitoring through the automation of the analysis of crowd video feeds and the delivery of real time metrics to facilitate intelligence-based decision-making in the management of crowds during Hajj.

II. Problem Definition

Current crowd-monitoring systems rely on manual observation, which does not scale well to thousands of cameras across multiple locations. Additionally, centralized video processing increases network load and raises privacy concerns. The main problem addressed in this project is the lack of an automated, scalable, and privacy-aware system for real-time crowd analysis.

HajjFlow aims to:

- Automatically estimate crowd size and pressure in real time.
- Reduce reliance on manual CCTV monitoring.
- Preserve privacy by processing video at the edge.

III. System Architecture

HajjFlow follows a distributed architecture consisting of edge devices, a central server, and a web-based dashboard.

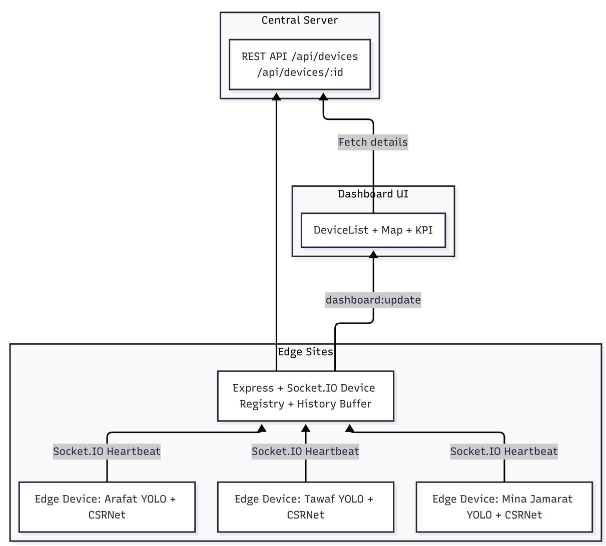

Fig. 1. High-level architecture of the HajjFlow system

Each edge device processes a video stream locally and sends periodic heartbeat messages containing crowd metrics to the central server. The server aggregates this data and updates the dashboard in real time.

IV. Computer Vision Methodology

A. YOLO-Based Person Detection

YOLO (You Only Look Once) is used for detecting individuals in scenes with low to medium crowd density. YOLO is a single-stage object detection model that offers high speed and good accuracy, making it suitable for edge deployment

[1], [2]. The system filters detections to the person class and counts visible individuals.

B. CSRNet for Dense Crowd Estimation

In very crowded scenes, individual detection becomes unreliable due to occlusion. To address this, CSRNet is used to estimate crowd density through density maps [3]. The total crowd count is obtained by integrating the density map.

C. Adaptive Switching Strategy

The system dynamically switches between YOLO and CSRNet based on crowd density indicators such as detection count and bounding box overlap. This ensures reliable estimation across different crowd conditions.

V. Edge Processing and Communication

All video analysis is performed on edge devices, and only numerical metrics are transmitted to the server using Socket.IO. This significantly reduces bandwidth usage and ensures that raw video data is not transmitted, supporting privacy-preserving design.

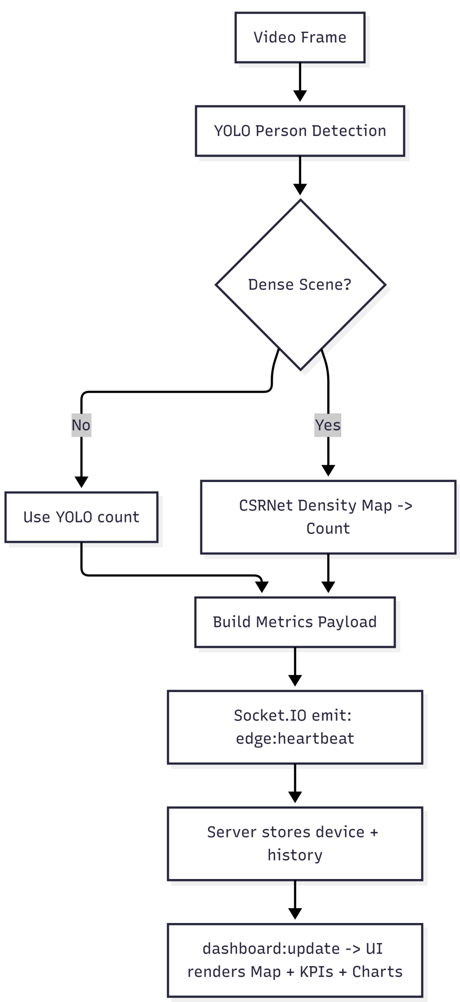

Fig. 2. Data flow from edge devices to the dashboard

VI. Results and Evaluation

The system was evaluated using prerecorded video footage representing three Hajj locations: Tawaf, Mina Jamarat, and Arafat. The evaluation focused on real-time performance, stability, and qualitative accuracy.

YOLO achieved real-time performance on CPU-based edge devices, with stable frame rates suitable for live monitoring. In dense scenes, CSRNet provided more consistent crowd estimates than person detection alone. The dashboard successfully displayed live device status, crowd metrics, and geographic locations with minimal latency.

Although ground-truth annotations were not available for quantitative accuracy measurement, qualitative results showed that the system correctly identified low, medium, and high crowd pressure scenarios. These results demonstrate the feasibility of the proposed approach for real-world deployment.

VII. Ethical and Practical Considerations

HajjFlow is designed with privacy in mind by ensuring that raw video data remains on edge devices. Only anonymized numerical metrics are transmitted to the server. Potential biases may arise due to camera placement, lighting conditions, or crowd composition, highlighting the need for careful validation before deployment.

From a practical perspective, the choice of edge hardware must balance computational capability, energy efficiency, and cost, especially when deploying deep learning models in large numbers.

VIII. Conclusion

This paper presented HajjFlow, an edge-based computer vision system for real-time crowd monitoring during Hajj. By combining YOLO-based person detection with CSRNet-based density estimation and deploying them within a scalable edge–server architecture, the system supports safer and more efficient crowd management. The project aligns with Saudi Vision 2030 and NSDAI goals by demonstrating the practical application of artificial intelligence for public safety and digital transformation.

References

[1] J. Redmon *et al.*, “You Only Look Once: Unified, Real-Time Object Detection,” in *Proc. CVPR*, 2016.

[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934, 2020.

[3] Y. Li, X. Zhang, and D. Chen, “CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes,” in *Proc. CVPR*, 2018.

[4] Kingdom of Saudi Arabia, “Saudi Vision 2030,” https://www.vision2030.gov.sa.

[5] Saudi Data & AI Authority, “National Strategy for Data and Artificial Intelligence (NSDAI),” https://www.sdaia.gov.sa.