Notification No.: 580/2025
Date of Award: 16/05/2025

Name of the Scholar: Sarthak Mishra
Name of the Supervisor: Dr. Suraiya Jabin
Name of the Department: Computer Science
Topic of Research: A Soft Biometric System to monitor Crowd.

Findings

Monitoring crowd behavior using CCTV cameras presents significant challenges, especially in identifying and analyzing anomalous activities within high-density environments. A critical component of crowd monitoring, pedestrian detection, encounters substantial difficulties in crowded settings due to issues like occlusion, variations in individual characteristics, and uncontrollable environmental factors such as lighting conditions and camera perspectives. This thesis proposed a deep learning-based solution by developing soft biometric feature extractors and integrating them with weakly supervised models enhanced by a temporal self-attention mechanism.

Leveraging deep learning techniques, advancements in pedestrian detection using models such as YOLOv4 have enabled faster and more accurate localization in crowded environments. Using transfer learning on the pre-trained YOLO model with the COCO dataset, this study applies Tiny YOLOv4 to the UCSD pedestrian detection dataset, achieving promising results in real-time crowd detection and tracking.

The work also explores video anomaly detection, where deep autoencoders analyze spatiotemporal features for identifying unusual activities in CCTV footage. A novel algorithm using autoencoders achieved competitive results on the UCSD Ped1 and Avenue datasets. Furthermore, a custom dataset, JMI-UCF, integrating Jamia University campus CCTV footage with the UCF Crime dataset, was developed. Leveraging pretrained 3D ConvNets and a Temporal Self-Attention (TSA) mechanism within a Multiple Instance Learning framework, the proposed system effectively detects anomalies with an impressive Average Precision of 96.26%. This research contributes to advancing real-time crowd surveillance through innovative methods for pedestrian detection, anomaly detection, and dataset creation, providing a robust foundation for ensuring safety and security in high-density environments like university campuses. The thesis is structured into 6 chapters which are summarized as follows:

Chapter 1 introduces critical area of crowd monitoring using computer vision techniques, particularly deep learning. It highlights the importance of analyzing crowd behavior for safety and understanding in public spaces, noting the challenges in high-density situations. The chapter outlines key components of an effective crowd monitoring system, including pedestrian detection, tracking, re-identification across multiple cameras, and the detection of anomalous events. It examines anomaly detection in crowd surveillance, categorizing methods by their learning approach and discussing feature extraction techniques like soft biometrics and 3D convolutional networks, ultimately aiming to design a real-time and accurate crowd surveillance system.

Chapter 2 offers a comprehensive survey of existing techniques in crowd monitoring and video anomaly detection. It details various publicly available benchmark datasets used for training and evaluating these systems. The review explores the evolution of AI in this field, from traditional computer vision methods to advanced deep learning architectures like CNNs and Vision Transformers, particularly focusing on pedestrian detection, tracking, and re-identification. Finally, it discusses recent trends in video anomaly detection along with an overview of feature extraction methods.

Chapter 3 explores the use of deep learning for pedestrian detection, highlighting the evolution from traditional methods to Convolutional Neural Networks. It details two prominent approaches: Faster R-CNN, which utilizes a region proposal network for efficient and accurate detection, and YOLO, a system known for its speed and global image processing. The chapter concludes by emphasizing the advancements and potential for further improvement in pedestrian detection systems using deep learning techniques.

Chapter 4 explores an unsupervised method for detecting anomalies in surveillance videos using deep autoencoders. The work introduces a novel algorithm that leverages both the spatial and temporal features of video data during training. The study demonstrates the effectiveness of their approach through quantitative metrics like AUC and visual regularity graphs that highlight anomalous video segments.

Chapter 5 introduces a novel, weakly supervised approach to video anomaly detection (VAD). The authors created a new dataset, JMI-UCF, by combining campus surveillance footage with specific categories from the UCF-Crime dataset to enhance training. They then designed and implemented models using the Multiple Instance Learning (MIL) framework and temporal self-attention mechanisms with feature extractors like C3D and I3D. The study explores the effectiveness of these methods on both the standard UCF-Crime dataset and the newly curated JMI-UCF dataset, emphasizing the importance of contextual data for improved anomaly detection in specific surveillance scenarios.

Chapter 6 concludes the thesis by discussing the effectiveness of proposed techniques. The chapter also discusses the the future scopes and limitations of the research work.