

CCTV Intrusion Detection Model Using Spatiotemporal Frequency Analysis

JeongHyeon Chang , Hyun Yoo

Contents Convergence Software Research Center, Kyonggi Universi,
Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, Republic of Korea
crime_tiger564@kyonggi.ac.kr, rhpa0916@kyonggi.ac.kr

Abstract: A CCTV-based intrusion detection system is a security technology that detects intrusions by analyzing real-time video feeds, commonly used in public spaces, businesses, and other critical infrastructures to safeguard assets and individuals. These systems are essential for ensuring security, as they provide continuous monitoring and alert capabilities. However, conventional systems often rely on simple rule-based models that lack adaptability to complex and dynamic environments, leading to frequent false positives and missed detections. For example, environmental factors such as lighting changes, shadows, or the movement of animals can trigger unnecessary alarms, creating inefficiencies and diminishing trust in the system's reliability. Additionally, these systems usually require intricate manual configurations, where users need to define specific detection rules for each camera or surveillance area. This can become particularly challenging in large-scale environments with multiple cameras, as it increases the risk of human error, slows down deployment, and limits the system's overall management efficiency.

To address these limitations, the proposed model acquires real-time streaming data from CCTV cameras using the RTSP protocol and utilizes the YOLO deep learning object detection algorithm to accurately detect objects such as people, vehicles, and other relevant entities within the video. YOLO's high-speed and precise object detection capabilities allow for more reliable identification of potential intrusions, even in real-time scenarios. Once the objects are detected, their positional data is fed into a spatiotemporal grid frequency analysis model, which examines the frequency of object occurrences over time and across different areas within the surveillance zone. Areas with low occurrence frequencies are automatically designated

as restricted zones using a threshold mechanism, reducing the need for manual input and constant monitoring. This automated approach not only lightens the load on users but also significantly enhances the system's ability to detect actual threats while minimizing false alerts. By integrating advanced deep learning techniques and automation, the model offers a more efficient, accurate, and user-friendly solution for modern security needs

Keywords: CCTV-based Intrusion Detection, YOLO, RTSP, Spatiotemporal Grid Frequency Analysis, Restricted Zones Detection

I. INTRODUCTION

Intrusion detection systems are security technologies primarily used in public spaces, businesses, and residential areas with the goal of protecting assets and people from various threats. Typically, CCTV cameras continuously record video, and the intrusion detection system monitors these video feeds to identify abnormal activities that match predefined rules or patterns. For instance, the system may trigger an alert when an individual enters a restricted area during specific times or crosses a designated boundary. However, conventional systems present several limitations. Rule-based systems are particularly prone to false positives and false negatives, especially in complex scenarios. For example, environmental factors such as changes in weather, shadows, or animal movements are often misinterpreted as actual intrusions, which exacerbates the problem. Additionally, processing large-scale video data from multiple cameras can overload memory and processing capabilities, thereby degrading system performance. This issue becomes more pronounced in large-scale environments where numerous CCTV cameras must be monitored and processed in real-time, leading to an increase

in the amount and complexity of data that the system must handle [1].

A significant challenge also lies in the complexity of the setup process and the cumbersome user interface. Many CCTV intrusion detection systems require users to define custom rules, which necessitates manual input of multiple parameters such as objects, locations, and time frames for each camera. Adjusting these rules to fit the environment can be time-consuming and difficult, especially for users unfamiliar with the technology.

Furthermore, managing multiple cameras simultaneously also poses a challenge for users. In large facilities with many installed cameras, monitoring all feeds individually is both inefficient and cumbersome. When alerts are triggered, it is often difficult to quickly identify which camera has issued the warning, complicating immediate responses to emergencies [2].

In this study, we propose an automated CCTV intrusion detection model that surpasses the limitations of conventional terminal AI CCTV systems by integrating advanced AI and deep learning-based analysis methods. Traditional systems often rely on predefined rules and manual configurations, which can be inflexible and prone to errors in dynamic or large-scale environments. By contrast, the proposed model enhances detection accuracy through the use of deep learning-based object detection, enabling the system to identify and classify objects such as people and vehicles with greater precision. Additionally, the model incorporates spatiotemporal frequency analysis to assess patterns of object occurrences over time and space, automatically designating restricted zones in areas where activity is infrequent. This frequency-based analysis ensures that the system adapts to changes in the surveillance environment without requiring constant manual input. As a result, the approach not only increases the reliability and accuracy of intrusion detection but also streamlines the configuration process by automating complex settings. This reduces the burden on users, making the system more efficient and user-friendly, especially in environments with multiple cameras or large monitoring areas.

II. RELATED RESEARCH

The proposed model acquires real-time streaming data from CCTV using the RTSP protocol and employs the YOLO deep learning object detection algorithm to quickly and accurately detect objects such as people and vehicles. Based on the positional information of detected objects, the model analyzes the frequency of object occurrences within the CCTV surveillance area, automatically designating low-frequency areas as restricted zones.

A. Data Acquisition via RTSP Streaming

The system begins by acquiring real-time streaming data directly from CCTV cameras using the Real-Time Streaming Protocol (RTSP). RTSP is a widely adopted and standardized protocol designed for the efficient transmission

of video and audio data over the network. This protocol is highly versatile, offering broad compatibility with various CCTV camera models and brands, making it ideal for diverse surveillance setups. Its flexible nature allows the system to integrate easily with existing infrastructure without the need for additional hardware or significant modifications. This capability to handle multiple camera streams concurrently is crucial in large-scale environments, such as airports, shopping malls, or industrial sites, where monitoring multiple locations at once is essential. Moreover, the use of RTSP enables seamless real-time streaming, ensuring that video feeds are processed with minimal latency, which is vital for time-sensitive applications like intrusion detection.

By continuously receiving and processing live data from multiple cameras, the system can promptly analyze the video feeds and respond to potential threats. The immediacy of real-time streaming allows for the rapid detection of unusual or unauthorized activities, such as a person entering a restricted zone or a vehicle moving into a prohibited area. This responsiveness is key to preventing security breaches or mitigating damage. In addition, the system's ability to process and analyze this data on-the-fly enables immediate alerts to security personnel, ensuring that corrective action can be taken before the situation escalates. With real-time data acquisition and processing, the model provides a robust and reliable solution for enhancing security, minimizing delays in detection, and improving overall operational efficiency [3].

B. Deep Learning-Based Object Detection

At the core of the proposed system is the YOLO (You Only Look Once) deep learning object detection algorithm, which is renowned for its ability to detect objects quickly and accurately. YOLO is a cutting-edge algorithm designed to achieve high performance in real-time video streams, making it ideal for environments where immediate detection is crucial. It can identify a wide range of objects, including people, vehicles, animals, and other potential threats, all while maintaining low latency. This real-time detection capability is essential for effective security systems, as any delay in identifying potential intrusions could lead to significant risks or damages.

What sets YOLO apart from traditional object detection algorithms is its ability to perform single-pass detection, meaning it processes the entire image in one go, rather than breaking it down into regions of interest and examining each separately. This results in significantly faster processing times, without compromising on accuracy. The system captures the positional information of detected objects as coordinates, which are then used to analyze and assess potential security threats. For example, the system can track the movement of an individual within a designated surveillance area or monitor the entry and exit of vehicles in restricted zones.

By integrating YOLO into the CCTV system, the model transcends basic perimeter-based intrusion detection,

allowing for a much more nuanced analysis. Instead of merely detecting the presence of an object, the system can now differentiate between object types, enabling a more sophisticated assessment of the situation. For instance, the system can distinguish between a harmless animal and a potential intruder or determine whether a vehicle is authorized to enter a specific zone. This level of detailed analysis significantly improves the precision of the system, reducing false positives and enhancing overall detection accuracy. Moreover, this approach opens new possibilities for advanced applications such as behavior analysis, crowd monitoring, and anomaly detection, further extending the system's capabilities beyond simple security monitoring [4], [5].

III. Spatiotemporal Grid Frequency Analysis Model

The detected objects are recorded based on their coordinates, and the occurrence frequency of each object within the CCTV surveillance area is analyzed. This process involves segmenting the video into a spatiotemporal grid, where object frequencies are calculated for each cell. If the frequency of object occurrences in a particular grid cell falls below a predefined threshold, that area is automatically designated as a restricted zone, indicating that intrusions are not permissible.

To implement this method, the entire surveillance area is divided into a grid with a fixed cell size. The grid is composed of two axes: time and space, with each grid cell representing a specific time interval and spatial region. For example, the entire area can be segmented into an $M \times N$ spatial grid, and each grid cell can be subdivided into specific time units to form the spatiotemporal grid. The frequency of objects detected in each grid cell is then calculated. The calculation process is outlined as follows:

- i. Recording Object Coordinates: The coordinates (x_i, y_i) and the occurrence time t_i of the detected object are recorded.
- ii. Grid Cell Assignment: Based on the object's coordinates, the grid cell (g_x, g_y) in which the object is located is determined, where g_x and g_y are the grid indices along the x-axis and y-axis, respectively.
- iii. Frequency Aggregation: The number of objects occurring within each grid cell during a specific time is aggregated. This allows for the calculation of the frequency F_{xy} for each grid cell.

$$F_{xy} = \frac{1}{T} \sum_{t_i \in T} \text{count}((x_i, y_i) \text{ in } (g_x, g_y) \text{ at time } t_i)$$

T : Represents the analysis period, and count refers to the number of object occurrences in a given time frame and grid cell. Subsequently, if the frequency of each grid cell falls below the threshold θ , the system automatically designates that area as a restricted zone. The threshold θ can be adjusted according to the system's requirements. The formula for automatic assignment based on the threshold is as follows:

$$\text{If } F_{xy} < \theta \text{ then } (g_x, g_y) \text{ is set as a restricted area}$$

In this process, areas with low frequency of activity are automatically designated as restricted zones, meaning that any object detected in these areas during the actual intrusion detection process is considered a potential security threat. For example, if a specific section of the CCTV surveillance area experiences little to no human or vehicle activity over an extended period, it is identified as a low-frequency zone. The system recognizes this zone as restricted, and any subsequent detection of objects within this area, such as an unauthorized person or vehicle, triggers an immediate alert to security personnel. This method enhances the ability of the system to focus on areas that are typically less trafficked, where activity is more likely to signal a breach or intrusion. By utilizing frequency analysis, the system can autonomously identify and classify high-risk areas without the need for manual configuration by the user. This automation significantly reduces the complexity of the setup process, making it more user-friendly and efficient, particularly in environments with a large number of surveillance cameras or extensive monitoring zones. Traditional systems often require users to manually define restricted zones and set up detailed detection rules, which can be time-consuming and prone to error. In contrast, the frequency-based approach allows the system to adapt to its environment dynamically, adjusting restricted zones based on real-world activity patterns.

This not only saves time for the user but also greatly improves the accuracy of intrusion detection by minimizing false positives and ensuring that alerts are only triggered in genuinely suspicious situations. Additionally, the system's ability to continuously update these restricted zones as activity patterns evolve over time ensures that security remains adaptive and responsive to changing conditions, further enhancing overall surveillance effectiveness.

IV. CONCLUSION

Traditional CCTV intrusion detection systems have the limitation of requiring users to manually configure surveillance zones and detection rules, which can be both inconvenient and restrictive. To address this issue, this study proposes an automated configuration method. The proposed approach integrates object detection from real-time streaming data with regional frequency analysis, enabling the system to automatically designate restricted zones and update detection rules. This approach allows for the automatic generation of optimal intrusion detection rules tailored to the CCTV installation environment without user intervention, thereby improving user experience and enhancing system operational efficiency.

Specifically, this method leverages a spatiotemporal grid structure to systematically analyze large-scale CCTV surveillance areas, reducing the complexity of the configuration process and improving detection accuracy through frequency-based automated zone designation. The method is flexible and adaptable to various environments and scenarios, ultimately enhancing the overall efficiency of CCTV-based intrusion detection systems, enabling more accurate and reliable detection.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.2020R1A6A1A03040583)

REFERENCES

- [1] S. Jain, "A framework for modeling and simulation for emergency response," in Proc. 2003 Winter Simulation Conf., vol. 1, IEEE, 2003.
- [2] S. S. A. Zaidi et al., "A survey of modern deep learning based object detection models," Digital Signal Processing, vol. 126, 103514, 2022.
- [3] A. Rao, "Real time streaming protocol (RTSP)," Internet-Draft, Nov. 1996.
- [4] N. Kehtarnavaz and M. N. Gamadia, "Real-time Image and Video Processing: From Research to Reality."
- [5] A. M. Durik et al., "The effects of hedges in persuasive arguments: A nuanced analysis of language," Journal of Language and Social Psychology, vol. 27, no. 3, pp. 217-234, 2008.

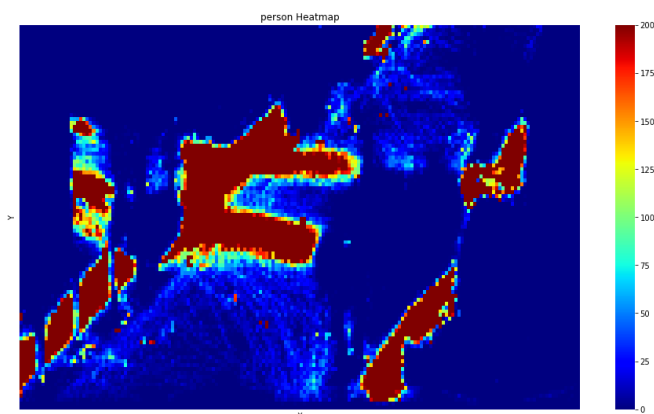


Figure 1. CCTV Footage Frequency Heatmap

The image above illustrates a frequency heatmap created by dividing street CCTV footage into grids according to a specific formula. Each grid represents a section of the street, and the heatmap visually depicts the frequency of occurrences within each area, helping to identify activity patterns or potential anomalies in the monitored space. This method of analysis provides a clear visual representation of how often different areas of the street are captured by CCTV, making it useful for surveillance and security applications.



Figure 2. Street Intersection for Frequency Analysis

Based on the uploaded image and the content we've discussed, here's a description that fits the context of frequency analysis in CCTV surveillance:

The image depicts a street intersection captured by a CCTV camera. This type of footage could be divided into grids for spatiotemporal frequency analysis, where each grid represents a specific section of the street. The frequency of occurrences (such as vehicles passing or pedestrians walking) within each grid could be tracked and visualized using a heatmap. This would allow for identifying patterns of activity or anomalies in the monitored space. The areas with higher frequencies of movement, such as major crosswalks or traffic lanes, could be highlighted, while less active areas might indicate potential zones for intrusion detection, depending on the context of the analysis.



JEONGHYEON CHANG received the Ph.D. degree in police administration from Kyonggi University, South Korea, in 2017. From 2019 to 2020, he was an Assistant Professor of science of international politics with the University of Halla. In 2020, he joined the Contents Convergence Software Research Institute, Kyonggi University, where he is currently a Research Professor. His research interests include crime prevention through active content, community policing, and data mining.



HYUN YOO received the B.S., M.S., and Ph.D. degrees from the Department of Computer and Information Engineering, Sangji University, South Korea, in 1999, 2011, and 2019, respectively. He was with Fujitsu and XCE (SK). He has been a Researcher with the Data Mining Laboratory, Kyonggi University, South Korea, where he is currently a Research Professor with the Contents Convergence Software Research Institute. His research interests include deep learning, artificial intelligent, big data mining, ambient intelligence, medical decision systems, healthcare, recommendation, and HCI.