I. INTRODUCTION

Street crime is a severe problem in Indonesia, especially in big cities. One of the street crimes that disturbs the community is the crime of motorcycle gangs. Motorcycle gangs are criminal groups engaged in various criminal activities, including aggression, property destruction, robbery, and even murder [1]. These groups often exhibit aggressive driving behaviors, such as speeding and reckless maneuvers, which pose significant risks to public safety and create a sense of fear and insecurity among residents.

Various prevention efforts have been made by authorities, such as the police and local governments, to prevent motorcycle gang crimes. One of the efforts made by the police is regular patrol activities, but this solution is inefficient due to limited time, personnel, and coverage area. Another solution local governments take is installing CCTV (Closed Circuit Television) to monitor the situation on the road. Some regions in Indonesia, especially big cities, have installed many CCTVs that are monitored through a centralized control center. However, this solution has several disadvantages. First, it requires many personnel to monitor hundreds of CCTV video broadcasts. Second, the possibility of detecting anomalies such as motorcycle gangs will decrease as the number of CCTV video broadcasts increases [2].

Many studies have been conducted to automate CCTV surveillance to detect crimes. One is research conducted by Atif Jan and Gul Muhammad Khan [2] with the Quasi-3D method to detect crimes in videos. The research aims to create a system to detect malicious video events using a modified CNN (Convolutional Neural Network) filter. Event detection in videos generally uses a 3-dimensional CNN as a feature extractor. However, this method has a drawback: the large number of parameters. In the article, the author separates a 2-dimensional CNN filter to learn spatial features on video frames with a CNN filter to learn temporal features between frames. After the features are extracted, the video will be classified based on its classes: normal, fighting, shooting, and vandalism.

In contrast to previous researchers who used supervised learning methods, Samir Bouindour et al. [3] used an unsupervised learning method for anomaly detection in videos. In that study, researchers only used normal event videos as training data. They used a modified pretrained 3D residual network to extract spatio-temporal features. They proposed a new method to detect outliers based on the output vector of the 3D residual network. This method can automatically select the vector of interest to distinguish between rare and anomalous events, thus reducing false alarms.

Similar to previous research, Fath U Min Ullah et al. [4] created a system to detect crime in videos by utilizing spatiotemporal features with 3D CNN. The system consists of three stages in detecting crime. First, it detects human objects in the input video using a lightweight CNN model such as MobileNet-SSD to obtain the human bounding box and discard useless frames. Next, the human-detected frames will be fed into a 3D CNN to obtain spatio-temporal features. Finally, these features will go to the softmax classifier to classify whether there is a crime in the video.

Furthermore, Sardar Waqar Khan et al [5] utilized CCTV to detect anomalous events on the highway in the form of traffic accidents. They used Deep Learning CNN to detect anomalous events in images from videos. Unlike previous studies that used the video's spatiotemporal (space and time) features, this study focuses on using spatial features in the image to detect anomalies. The system's input is a video split frame by frame into images. Then, the image will be classified using the CNN model. If an accident is detected, the system will send an accident notification to the authorized officer.

While these methods effectively detect individual actions, they must be improved when applied to more complex scenarios like motorcycle gang activities. In such cases, the collective behavior of multiple motorcyclists creates unique movement patterns that cannot be captured by focusing on a single actor. Therefore, a new approach is needed—one that can learn and analyze the movement patterns of all motorcyclists to distinguish between motorcycle gang activities and normal motorcyclist behavior.

In this study, we propose a new approach to learning the collective movement of motorcyclists. The proposed method for detecting motorcycle gang activity consists of three key stages. First, object detection and tracking algorithms are combined to detect and track motorcycles in video footage. A variety of studies on the topic of object detection have been conducted, including those referenced in [6], [7], [8]. A combination of object detection and object tracking algorithms has been the subject of study in [9]. The second step involves extracting and mapping each tracked motorcycle's centroid coordinates based on the tracking IDs assigned to them. Finally, these motorcycles' motion patterns are examined and classified to separate normal behavior from gang-related actions. This multi-step approach provides a structured and precise classification process, enhancing the ability to detect motorcycle gangs in video footage more accurately.

The novelty of this research lies in several innovative components. (1) Integrating YOLOv9 with ByteTrack facilitates the robust detection and consistent tracking of motorcycles in real-world video footage. (2) The introduction of the Centrogen algorithm maps the centroid coordinates of motorcycles into structured 30x30 matrices, enabling a systematic representation of their movement across frames. (3) The design of a classification framework that leverages these matrices distinguishes between normal and gang-related motorcycle movements. These innovations collectively enhance the accuracy and scalability of motorcycle gang detection in surveillance applications.