

High-Accuracy Vehicle Detection in Different Traffic Densities Using Improved Gaussian Mixture Model with Cuckoo Search Optimization

Nor Afiqah Mohd Aris, Siti Suhana Jamaian

Department of Mathematics and Statistics-Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub, Pagoh, 84600, Malaysia

Abstract—Background subtraction plays a critical role in computer vision, particularly in vehicle detection and tracking. Traditional Gaussian Mixture Models (GMM) face limitations in dynamic traffic scenarios, leading to inaccuracies. This study proposes an Improved GMM with adaptive time-varying learning rates, exponential decay, and outlier processing to enhance performance across light, moderate, and heavy traffic densities. The model's parameters are automatically optimized using the Cuckoo Search algorithm, improving adaptability to varying environmental conditions. Validated on the ChangeDetection.net 2014 dataset, the Improved GMM achieves superior precision, recall, and F-measure compared to existing methods. Its consistent performance across diverse traffic scenarios highlights its effectiveness for real-time traffic flow analysis and vehicle detection applications.

Keywords—Gaussian mixture model; vehicle detection; adaptive time-varying learning rate; exponential decay; outlier processing; cuckoo search optimization

I. INTRODUCTION

In recent years, the significance of vehicle detection and tracking systems has increased, driven by the growing demand for efficient and intelligent transportation systems. These systems play a pivotal role in diverse applications such as traffic management, accident prevention, and autonomous vehicle navigation [1-5]. To meet these demands, accurate and dependable Background Subtraction (BS) methods are essential in handling the complexities posed by dynamic and different traffic scenarios. Researchers have proposed many methods to study vehicle detection, such as traditional computer vision, machine learning, deep learning, motion-based, radar-based, and fusion techniques. This research specifically focuses on traditional computer vision methods: background subtraction in terms of mathematical contributions.

The BS method is a technique used for object detection. It involves segmenting the foreground from the background scene by generating a binary mask that identifies moving objects. The core principle of the BS method is to compute the difference between the current frame and a reference frame (background image). Thresholding techniques are then applied to classify the segmented pixels as either foreground or background. This process effectively isolates moving objects from the stationary background.

BS is a pivotal component in detection, tracking, and scene understanding, has been extensively addressed through GMM. GMM, known for their efficacy, are widely utilized to model complex and multi-modal background scenes by capturing statistical distributions of pixel intensities over time [6-8]. However, the application of traditional GMMs encounters notable challenges in the domain of vehicle detection, particularly when faced with different traffic densities. These challenges stem from the inherent limitations of constant learning rates in traditional GMMs, which are unable to adjust dynamically to the varying characteristics of the data. Traffic density and vehicle movement patterns can vary significantly over time and across environments. In such scenarios, a fixed learning rate often proves suboptimal, leading to issues such as slow convergence or convergence to suboptimal solutions, thereby impacting detection accuracy.

Moreover, traditional GMMs treat all observations equally, including outliers, which can distort the underlying data distribution and reduce the precision of the segmentation. They also assume that pixels do not closely match the mean belonging to the same statistical cluster, which may not be accurate in highly dynamic environments. These shortcomings hinder the ability of traditional models to adapt effectively to rapid changes in traffic conditions, such as those encountered in heavy or fluctuating traffic densities.

To address these limitations, this study proposes an Improved Gaussian Mixture Model (Improved GMM) that builds upon the strengths of traditional GMMs while introducing key enhancements. The Improved GMM incorporates an adaptive time-varying learning rate, which allows the model to dynamically adjust its parameters based on the characteristics of the current data. This adaptability improves performance across different traffic densities by better accommodating environmental changes.

Additionally, the model introduces exponential decay, which emphasizes pixels closer to the mean, enhancing the model's ability to distinguish between objects and background elements with higher precision. Outlier processing is also incorporated to control the influence of new covariance update observations, ensuring robustness against noisy data and outliers. These modifications collectively enable the Improved GMM to handle the complexities of vehicle detection under varying traffic densities.

Finally, to further enhance the robustness and adaptability of the Improved GMM, the Cuckoo Search (CS) optimization technique is employed for automatic parameter tuning. Inspired by the breeding behavior of cuckoos, this metaheuristic algorithm intelligently selects optimal values for critical parameters, such as the number of Gaussian components and learning rates, by exploring the parameter space efficiently. Unlike manual tuning, CS optimization dynamically adapts to the complexity of different traffic scenes, ensuring that the Improved GMM consistently delivers high detection accuracy across diverse environmental conditions with minimal human intervention.

Traffic density affects the dynamics of the background scene. In light traffic conditions, background updates may occur less frequently as there are fewer changes to the background. In contrast, the background may change rapidly in heavy traffic conditions due to the movement of multiple vehicles. Understanding and adapting to these variations in background dynamics is essential for accurate background subtraction [9-10].

The focus extends beyond the general challenges of vehicle detection to performance under different traffic densities—light, moderate, and heavy traffic conditions. The classification can be defined by the number of vehicles per square foot. Light traffic scenarios may involve less than three vehicles distributed per 500 square feet; moderate traffic represents a balance of vehicles, which is less than five vehicles distributed per 500 square feet, while heavy traffic introduces challenges, such as more than six vehicles distributed per 500 square feet. These variations necessitate developing an adaptive method to effectively address these issues while maintaining high detection accuracy and computational efficiency.

The structure of this paper is organized as follows: Section II reviews related works, providing an overview of existing approaches and background subtraction enhancements. Section III comprehensively discusses GMM and its applications in background subtraction. Section IV details the proposed Improved GMM, outlining modifications and introducing key elements. Section V describes the experimental setup and datasets utilized for evaluation, while results and discussion is given in Section VI and finally, the paper is concluded in Section VII.

II. RELATED WORKS

In the realm of computer vision, the study of background subtraction has been both significant and extensively explored. This led to the development and presentation of numerous methods and techniques to overcome diverse challenges, particularly in the context of detecting vehicles within dynamic traffic scenarios. The GMM has been widely adopted for background subtraction in computer vision applications, particularly in the vehicle detection field [7]. Initially, Friedman and Russel introduced the GMMs for background subtraction [11], while Stauffer and Grimson subsequently developed effective modified equations [6]. Numerous researchers also proposed additional modifications and improvements on the original model to enhance its performance in various traffic scenarios. Hence, this section

reviews several key developments and recent field advancements, focusing on an overview of the existing GMM methods and their improvements in background subtraction. Stauffer and Grimson denoted the GMM as a background subtraction method, which gained significant popularity due to its ability to model complex and multi-modal background scenes [6]. The study successfully achieved the objectives by capturing the statistical distribution of pixel intensities over time. Consequently, numerous researchers proposed improvements and modifications to the traditional GMM.

Zivkovic developed an adaptive GMM with a configurable number of Gaussian components, producing improved model adaptations to changing conditions [7]. Meanwhile, Zuo et al. designed an enhanced method for noise interruption for the traditional GMM [12]. The study incorporated several techniques to improve performance, including image block averaging, wavelet semi-thresholding, and adaptive background updating. Thus, the method effectively eliminated noise issues and enhanced the detection performance of the moving targets. The study also utilized an adaptive background update during the background updating phase, resulting in more accurate detection results. Another study by Lin and Chen discovered a novel method for recognizing moving objects that integrated GMM with visual saliency maps [13]. The approach effectively overcame the challenges caused by shadow situations while producing stable detection results, which transformed each image frame to the L^* , a^* , and b^* colour spaces. A Gaussian filter was then utilized to smooth the L^* , a^* , and b^* channels, eliminating small texture features and noise. The saliency maps were estimated for each channel and linearly merged to generate a comprehensive saliency map, which was combined with the foregrounds to obtain the moving objects.

Meanwhile, Zivkovic and Heijden present two efficient adaptive density estimation methods for background subtraction in video surveillance systems. The first method is based on a GMM and uses recursive equations to update model parameters and select appropriate components for each pixel. The second method is a nonparametric kernel-based approach that adapts to changes in the scene by updating the training data set. The performance of both methods is evaluated and compared to other algorithms. The results show that the nonparametric method outperforms the GMM approach in terms of accuracy but at the cost of increased processing time. This research provides valuable insights into the challenges of background subtraction and offers practical solutions for real-world applications [14]. A study by Varadarajan et al. proposes a new approach to modeling and subtracting backgrounds effectively in scenes with complex dynamic textures. The proposed method considers the spatial relationship between pixels, modeling regions as mixture distributions rather than individual pixels. In this research, the researchers derive novel online update equations using expectation maximization (EM) for modeling scenes containing dynamic textures. The effectiveness of the proposed algorithm is experimentally verified on various video sequences and compared with other well-known background subtraction algorithms. The results show that the proposed algorithm performs better than most algorithms and produces comparable results to ViBe, one of

the best background subtraction algorithms currently in the literature [15].

Another study by Cioppa et al. introduces a novel algorithm called Real-Time Semantic Background Subtraction (RT-SBS), which combines real-time background subtraction with high-quality semantic information for improved performance. The algorithm addresses the limitations of traditional background subtraction methods by leveraging semantic information provided at a slower pace and for some pixels. RT-SBS reuses previous semantic information by integrating a change detection algorithm during the decision process, ensuring real-time applicability while maintaining performance comparable to SBS. This work advances real-time background subtraction algorithms, particularly in scenarios with dynamic backgrounds, illumination changes, and moving objects [16].

Işık et al. proposed a novel method for foreground or background extraction in videos, specifically designed to address challenges posed by dynamic backgrounds. The Common Vector Approach for Background Subtraction (CVABS) leverages the Common Vector Approach (CVA) obtained through Gram-Schmidt Orthogonalization to achieve accurate background modeling. By treating background modeling as a spatiotemporal classification problem, the algorithm computes the common vector of frames to acquire the background model, enabling effective foreground object detection. The method incorporates a self-learning feedback mechanism to mitigate the impact of dynamic scenes and illumination changes on foreground detection accuracy. Experimental evaluations on diverse, dynamic backgrounds demonstrate the effectiveness of CVABS, positioning it as a competitive solution in the field of moving object segmentation [17].

In recent years, researchers also emphasized optimizing the GMM framework for specific applications, such as vehicle detection in traffic scenarios. A study by Zhang et al. described GMM with Confidence Measurement (GMMCM) as a potential solution [18]. The study addressed the susceptibility of background subtraction models towards contamination by slowly moving or temporarily stopping vehicles. Furthermore, the GMMCM incorporated a Confidence Measurement (CM) technique, assigning trust values to each pixel in the background model. This method quantified the current reliability of background pixels, and the design was developed to balance the dynamic changes in brightness and background (resolving contamination challenges) in complex urban traffic scenes. Consequently, this method was successful through a self-adaptive learning rate, which ensured the background model remained accurate. Another study by Lima et al. included a method for estimating the region-specific thresholds using a feedback step [19]. The approach employed spatial analysis to select an appropriate threshold for each region, which was utilized for pixel classification. A filtering technique was applied to the segmentation before the threshold estimate to address classification errors. This filtering process eliminated disturbances and consolidated the entire area into a cohesive unit. During the feedback phase, segmentation corrections estimated the thresholds for subsequent iterations. Notably, the filtering stage focused on correcting foreground errors, significantly enhancing the vehicle areas. This

recommended strategy facilitated the segmentation of previously segmented regions and resembled a first-order Markov chain estimate of the threshold.

In a study by Agrawal and Natu, a novel approach was developed by combining GMM with blob analysis, including labelling and morphological operations, to enhance the accuracy of foreground detection [20]. The model computed the difference between the reference frame BMG (x, y) and the current frame while applying a threshold to isolate the region of interest. In constructing the foreground model, a threshold value was selected for each pixel, which was determined using the standard deviation. A study by Luo et al. summarised a motion detection method considering spatial variation in image thresholds [21]. The approach required calculating the projected motion size under different image regions, established using a mapping correlation between the geometric motion features and the appropriate enclosing rectangle (BLOB) level in the spatial domain. This discovery enabled an adaptive threshold for each motion, effectively removing unwanted noise during motion detection.

Chen and Ellis employed a multi-dimensional Gaussian Kernel Density Transform (MDGKT) pre-processor to reduce noise in the spectral, temporal, and spatial domains [22]. This pre-processor applied spatial and temporal smoothing to each spectral component using a multivariate kernel, regarded as the product of two radially symmetric kernels. The MDGKT was a crucial component in improving the reliability of the GMM. Thus, the time interval and resolution of the GMM were changed by modifying the size of the kernel through a pair of bandwidth parameters. Kalti and Mahjoub designed a unique approach that incorporated a fuzzy distance into the Expectation-Maximisation (EM) and Adaptive Distance-based Fuzzy-C-Means (ADFCM) algorithms [23]. The pixel characterization in the study was based on two factors: the inherent attributes of the pixel and the characteristics of its surrounding neighbourhood. The classification was then measured using an adaptive distance that preferred one of the attributes concerning the pixel spatial location within the image. Another study by Wei and Zheng studied a method that calculated the L2 norm between the GMMs to measure the similarity corresponding to two pixels [24]. The study recorded the grayscale information of the pixel and the feature abundance in the local image region. Compared to individual pixels based on their differences, higher accurate pixel intensity measurements and information variation in the surrounding region were obtained. This similarity-based approach enhanced the performance of image-denoising models and preserved the detailed information in the image. Likewise, Chen and Ellis discussed an innovative approach that addressed the global illumination change concern in background model adaptation [22]. The study applied a revised adaptive strategy within the iterative learning process of the Zivkovic-Heijden GMM (ZHGM). This method was implemented by integrating a modified adaptive schedule into an existing filtering system, yielding superior performance than previous approaches (particularly in scenarios involving sudden illumination changes).

Martins et al. designed a novel classification mechanism that combined colour space discrimination, hysteresis, and

dynamic learning rate to address sudden illumination changes in the background model [25]. Each channel element (L*, a*, and b*) was analyzed individually, and the decisions obtained from each channel were merged using the AND rule, producing superior results than majority voting. This approach ensured a faster model and slower adaptations in dynamic and static regions. A higher learning rate (α_{UBG}) was applied if the pixel classification transitioned from foreground to background. Therefore, this mechanism promoted rapid adaptation when the background reappeared, effectively preventing the phantom image from emergence.

Regarding mathematical contributions, Su introduced a GMM with a data model optimization approach to address the adapting challenge of light transitions [26]. The initial step in the process included gradient picture calculation of the video stream using the Scar Operator. Subsequently, the RGB values and gradients were integrated, and noisy movement areas were eliminated using various techniques (combining the remaining sites). The two model outputs were compared to determine the final makeup area in mitigating incorrect diagnosis risk. Thus, the results demonstrated that the approach enhanced the detection process accuracy by minimizing the erroneously detected area occurrences caused by sudden illumination changes.

The advancements in GMM-based background subtraction have significantly addressed noise, dynamic textures, and illumination changes. Approaches such as RT-SBS, CVABS, and adaptive density estimation have demonstrated success in handling specific scenarios, including dynamic backgrounds and complex traffic environments. However, these methods often rely on manually tuned or fixed parameters, hindering their adaptability to fluctuating traffic densities and varying environmental conditions. This study bridges these gaps by proposing an Improved GMM that incorporates adaptive time-varying learning rates, exponential decay, and robust outlier processing, supported by CS Optimization for automatic parameter tuning. This ensures consistent and robust performance across light, moderate, and heavy traffic scenarios, contributing to the development of efficient vehicle detection systems.

III. GAUSSIAN MIXTURE MODEL IN VEHICLE DETECTION

GMM is a Mixture of Gaussians (MoG), a prominent strategy for background subtraction methods in computer vision applications. A study by Stauffer et al. discovered this strategy based on a parametric model in handling multiple modes within the pixel values [6]. The study implied that the background and foreground distributions for the GMM were Gaussian, in which the background area was more visible and exhibited smaller variances than the foreground. This assumption enabled the GMM to effectively manage slow-lighting changes and -moving objects, periodic motion, long-term scene changes, and camera noise. Conversely, the GMM was only used for its computational efficiency and excellent performance in numerous applications, as the previous assumption was not always true.

The GMM aims to construct a background model for each pixel to the time-domain distribution of pixel values in a video sequence. This model represents the weighted sum of a finite

number of Gaussian functions, which describes the multi-peak state of pixels while being suitable for complex background models (light gradients and swaying trees). In the GMM, Gaussian components with large weights represent the background, while those with small weights represent the foreground. Generally, a new pixel is part of the background if it correspondingly matches the Gaussian model. Otherwise, the pixel is treated as a foreground pixel if it does not match a Gaussian model (or match a Gaussian model with only a small weight). The efficacy of GMM has prompted various improvements and extensions in the field, which has become a widespread practice for background extraction in computer vision applications [7, 10, 26, 27, 28]. Hence, the application of GMM in vehicle detection can be expressed as in Eq. (1), where the weighted sum of K Gaussian distributions times the Gaussian component.

$$f(x_t) = \sum_{k=1}^K \Pi_{k,t} \cdot \Phi(x_t, \mu_{k,t}, \sigma_{k,t}) \quad (1)$$

where x_t is the pixel value; $\Phi(x_t, \mu_{k,t}, \sigma_{k,t})$ is the Gaussian component density with mean $\mu_{k,t}$ with covariance matrix $\sigma_{k,t}$; $\Pi_{k,t}$ is the weight associated with the k^{th} Gaussian component. Subsequently, $\Phi(x_t, \mu_{k,t}, \sigma_{k,t})$ is formulated as:

$$\Phi(x_t, \mu_{k,t}, \sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sigma_{k,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (x_t - \mu_{k,t})} \quad (2)$$

where n is the dimension of the pixel intensity. The covariance matrix is also assumed as $\sigma_{k,t} = \sigma_{k,t}^2 I$. Each new pixel value, x_t is compared with each of the existing K Gaussian distributions. A pixel is considered to match a Gaussian distribution if its value falls within a range of 2.5 standard deviations from the mean of that distribution where the matching condition is $|x_t - \mu_{k,t-1}| \leq 2.5\sigma_{k,t-1}$. The classification process involves categorizing a pixel as background if it matches with the Gaussian distribution identified as background, and as foreground if it matches with the Gaussian distribution identified as foreground. In cases where the pixel does not match with any of the K Gaussians, it is classified as foreground. This process results in the creation of a binary mask. A new Gaussian distribution is added if $k < K$, while the Gaussian distribution is replaced with the lowest priority $k = K$ ($\sigma_{k,t}^2 = \Pi_{k,t} / \sigma_{k,t}$) if $k = K$. The weights of every Gaussian distribution must be updated for the next foreground detection,

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha\psi_{k,t} \quad (3)$$

where ψ is the indicator function and α is the constant learning rate. The mean and variance that do not find a match remain unchanged. However, for the component that does match, its mean and variance are updated according to the following criteria:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_t \quad (4)$$

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (5)$$

where $\beta = \alpha \cdot \Phi(x_t, \mu_{k,t}, \sigma_{k,t})$. If the k^{th} Gaussian distribution matches x_t , then $\psi = 1$. Otherwise, $\psi = 0$ if the k^{th} Gaussian distribution does not match with x_t . The Gaussian distribution weights are then normalized after being modified. The K Gaussian distribution for each pixel is described after the modification process as:

$$B = \operatorname{argmin} \left(\sum_{k=1}^b \Pi_k > th \right) \quad (6)$$

where th is the threshold. Based on the ratio (Π/σ) , these distributions are listed by priority order, beginning with the B Gaussian distribution. Subsequently, a continuous comparison of the x_t and B Gaussian distribution is performed. The pixel is considered a background point if the x_t distribution matches any preceding B Gaussian distribution points. Alternatively, the pixel is regarded as a foreground point if it does not match, and the moving object detection is considered complete.

IV. PROPOSED IMPROVED GMM

The proposed Improved GMM model is fixed as in Eq. (1) and the Gaussian component in Eq. (2), but there are some modifications in Eq. (3) to Eq. (5), which are the updating parameters of the Gaussian component. This section presents the GMM modifications for achieving high accuracy in vehicle detection across various traffic densities, which addresses the traditional GMM limitations by introducing an adaptive time-varying learning rate, exponential decay, and outlier processing. Traditional GMMs have limitations, such as fixed learning rates, sensitivity to outliers, and difficulties in distinguishing between closely spaced objects. Therefore, these enhancements allowed the Improved GMM to effectively capture the dynamic nature of traffic density variations and improve detection accuracy. The suggested Improved GMM was founded on the improvements to the mean, covariance, and weight equations presented by Stauffer and Grimson [6]. The upgrades are as follows:

$$\mu_{k,t} = (1 - \beta(t))\mu_{k,t-1} + \beta(t)x_t \cdot e^{-\lambda d_t} \quad (7)$$

$$\sigma_{k,t}^2 = (1 - \beta(t))\sigma_{k,t-1}^2 + \beta(t)(1 - \gamma)(1 - e^{-2\lambda d_t})(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (8)$$

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha e^{-\lambda d_t} \quad (9)$$

where $\beta(t) = c/(c + t)$.

A. Adaptive Time-Varying Learning Rate

Traditional GMMs use a constant learning rate, which struggles to adapt to changing traffic densities. For example, in heavy traffic, where vehicles move closely together, or in light traffic, where vehicles are sparse, a fixed learning rate may result in slow adaptation or inaccurate background modeling. An adaptive time-varying learning rate controls the weights assigned to the newly arriving data samples [30]. This suggestion enables the algorithm to quickly adapt to traffic flow changes and detect vehicles more accurately by considering the distance between the pixel and the current means. This modification is mathematically represented by Theorem 1.

Theorem 1. The $\beta(t)$ is derived using the Robbins-Monro stochastic approximation method, which involves solving the recursive equation as:

$$\beta(t) = c/(c + t) \quad (10)$$

where c is a constant that controls the learning rate.

Proof. The Robbins-Monro stochastic approximation method is an iterative algorithm to solve root-finding issues for non-linear equations in form $f(x) = 0$, which is based on stochastic gradient descent [31, 32, 33]. The Robbins-Monro conditions are satisfied to demonstrate the validity of the update rule in Eq. (10). This validity ensures the convergence of the stochastic approximation method, which two main criteria of the Robbins-Monro conditions are as follows:

Condition 1: The sum of the learning rates $[\sum_t \beta(t)]$ should diverge and $\sum_t \beta(t) = \infty$. By evaluating the summation, a telescoping series is expressed as:

$$\sum_t \beta(t) = \frac{c}{c+1} + \frac{c}{c+2} + \frac{c}{c+3} + \dots + \frac{c}{c+t} \quad (11)$$

When the terms are rearranged, they can be written as:

$$\sum_t \beta(t) = c \left[\frac{1}{c+1} + \frac{1}{c+2} + \frac{1}{c+3} + \dots + \frac{1}{c+t} \right] \quad (12)$$

Since Eq. (12) diverges, it can reach infinity as t approaches infinity. Therefore, $\sum_t \beta(t)$ also diverges, satisfying Condition 1.

Condition 2: The $\sum_t \beta^2(t)$ should converge and $\sum_t \beta^2(t) < \infty$. By expanding and simplifying the expression, an equation is formulated as:

$$\sum_t \beta^2(t) = \frac{c^2}{(c+1)^2} + \frac{c^2}{(c+2)^2} + \frac{c^2}{(c+3)^2} + \dots + \frac{c^2}{(c+t)^2} \quad (13)$$

As in Eq. (13) converges, a finite sum is demonstrated as t approaches infinity. Hence, $\sum_t \beta^2(t)$ also converges, satisfying Condition 2. When both conditions are satisfied, the updated rule in (10) is proven valid within the context of the Robbins-Monro stochastic approximation method [34]. This updated rule ensures the learning algorithm convergence as the iteration or t approaches infinity. This strategy provides more weight to the newer data pixels while maintaining a certain importance level for the past data pixels. The constant (c) in the Robbins-Monro stochastic approximation method controls the learning rate and should be chosen based on the data characteristics and the specific application [35].

The c parameter value is selected based on a priori data knowledge, such as possible value ranges for the model parameters and the data distribution. Notably, the c value affects the performance of the algorithm, which selecting the incorrect value leads to slow convergence or instability. The value of the iteration or t is typically set to increment by one with each iteration. The t value is also interpreted as the number of algorithm iterations or observations analyzed, which the t initial value and the growth rate impact the convergence speed and algorithm stability. If the t initial value is too small, the step size can be excessively large, causing instability and

overestimating the optimal solution. Similarly, if the t initial value is too high, the step size can be extremely small, leading to slow convergence and the possibility of becoming trapped with a suboptimal solution. The t growth rate also affects the convergence speed and algorithm stability. A rapid increase in t promotes faster convergence, which leads to instability and overestimation. Alternatively, a slower growth in t produce stable behaviour, which leads to slow convergence.

B. Exponential Decay

Exponential decay improves the model's ability to handle dynamic objects and lighting changes by giving more weight to recent pixels closer to the mean. For example, when vehicles move closer to the camera, their pixels influence foreground detection more significantly. The parameter λ adjusts the contributions of each current pixel (mean, covariance, and weight parameters) to the Improved GMM. When this adjustment is incorporated, the vehicle detection precision increases to the distance between the current pixel and the current mean. The $\mu_{k,t-1}$ is the previous mean, x_t is the current pixel, and d_t is the Euclidean distance between x_t and the previous mean $\mu_{k,t-1}$. Furthermore, the introduction of the exponential decay factor ($e^{-\lambda d_t}$) from Eq. (7) to Eq. (9) ensures that the contribution of each pixel decreases as its distance from the current mean increases [36]. This characteristic is consistent with the notion that pixels closer to the mean influence the parameter changes more than those further away [37]. The additional term $(1 - e^{-2\lambda d_t})$ adjusts the contribution of each pixel to the covariance parameter based on its distance from the current mean. From Eq. (7) to Eq. (9), $\Pi_{k,t-1}$ represents the previous weight, and $e^{-\lambda d_t}$ modifies the contribution of each data point to the weight parameter with respect to its distance from the current mean. This modification assures that the contribution of each data point to the weight reduces as its distance from the current mean increases [38]. The updated parameter equation is explained in the Theorem 2:

Theorem 2. Consider $e^{-\lambda d_t}$ to be the exponential decay factor introduced from the Improved GMM ((7) to (9)) for vehicle detection in real-time traffic flow analysis. Based on the distance from the current mean, this exponential decay factor modifies the contribution of each current pixel to the mean, covariance, and weight parameters. Therefore, the equations are:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_t \cdot e^{-\lambda d_t} \quad (14)$$

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(1 - e^{-2\lambda d_t})(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (15)$$

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha e^{-\lambda d_t} \quad (16)$$

Proof. Consider two pixels (x_{t1} and x_{t2}), where d_{t1} is the distance between x_{t1} and the mean. In addition, d_{t2} is the distance between x_{t2} and the mean. Assume $d_{t1} < d_{t2}$. Hence, x_{t2} and d_{t2} are substituted into Equation 14 produces a new equation:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_{t2} \cdot e^{-\lambda d_{t2}} \quad (17)$$

Since $d_{t1} < d_{t2}$, the $e^{-\lambda d_{t1}}$ is greater than $e^{-\lambda d_{t2}}$. The x_{t1} contribution to the mean parameter is higher than x_{t2} . Similar

reason is extended to Eq. (15) and Eq. (16) to demonstrate that $e^{-\lambda d_{t1}}$ influences the covariance and weight parameters (in a manner that decreases the pixel contribution as its distance from the mean rises) [39].

The contribution of each pixel to the mean parameter decreases as the pixel moves away from the current mean. This outcome prevents outliers or noisy data points from significantly affecting the estimated mean [40]. If a pixel is far from the current mean, it may not accurately represent the underlying distribution, so its contribution to the mean should be downweighed. In vehicle detection, a pixel corresponds to a vehicle far from the current mean is an outlier or an erroneous detection caused by noise in the sensor data. When downweighing the contribution of each pixel to the mean, the algorithm is more robust to such outliers and minimizes the likelihood of false positives or misclassifications. Therefore, adjusting the contribution of each data point to the mean parameter based on its distance from the current mean improves the accuracy and robustness of the GMM algorithm for vehicle detection in real-time traffic flow analysis.

C. Outlier Processing

Traditional GMMs treat all data points equally, including outliers, which can distort the background model. For instance, sudden reflections, shadows, or noisy pixels may lead to false detections. To address this, a robustness parameter (γ) is added to control how much the model updates with new data [41]. This modification permits the influence control of new observations on covariance matrix updates [42]. The γ in the covariance updated equation of the GMM improves the robustness of the model against outliers and noisy data. This new equation is expressed as:

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(1 - \gamma)(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (18)$$

The updated equation treats all observations in traditional GMM, including the outliers. Hence, the covariance matrix adapts to the outliers and incorporates their influence, potentially leading to distorted estimates of the underlying data distribution. On the contrary, the Improved GMM presented with γ allows the influence control of the new covariance update observations. When adjusting γ value, the weight provided to outlier-like observations is reduced, effectively downplaying their impact on the covariance estimation. When γ is closer to 1, the impact of new observations is reduced, making the model more robust to outliers. Robust estimators mitigate the influence of outliers or departures from model assumptions, providing more reliable parameter estimates [43].

The model responsivity is controlled by incorporating γ into the covariance update equation. This control mechanism balances integrating new information and protecting against outliers. When γ is adjusted, the weight given to outlier-like observations is reduced, allowing the model to focus more on reliable and representative data points [44]. Extensive experiments on real-world datasets were investigated in this study to assess the effectiveness of the proposed modification. The performance of the traditional GMM formulation was then compared with the modified version, incorporating an adaptive

time-varying learning rate, exponential decay, and outlier process.

V. COMPUTATIONAL SETUP

A. Algorithm

The discussion above outlines the computational algorithm for Improved GMM by introducing an adaptive time-varying learning rate, exponential decay, and outlier processing. The steps are similar to traditional GMM, except for steps (4) and steps (5), where there are modifications to the updating parameters means, variance, and weight. The algorithm can be summarized as:

Algorithm 1: Improved GMM

Initialize $K, k, \alpha, \lambda, \gamma, d$ and c

- Step 1 Obtain the current pixel value x_t from video frames.
 - Step 2 Compare the current pixel value with k existing Gaussian distributions, with a matching condition $|x_t - \mu_{k,t-1}| \leq 2.5\sigma_{k,t-1}$. A match is identified if the pixel value falls within 2.5 standard deviations of the distribution.
 - Step 3 If the matching condition is satisfied, classify the pixel values as background. Otherwise, consider the pixel value part of the foreground. If $k < K$, add a new Gaussian distribution; otherwise, replace the Gaussian distribution with the lowest priority $k = K$.
 - Step 4 Update expressions in (7) and (8) if mean and standard deviation do not match.
 - Step 5 After match inspection, update (9) and normalize it after modification.
 - Step 6 List distributions by priority order based on the ratio Π/σ , starting with the B Gaussian distribution.
 - Step 7 Conduct a continuous comparison of x_t and (6). Classify the pixel as a background point if it matches any preceding B Gaussian distribution points; otherwise, consider it a foreground point, indicating the completion of moving object detection.
-

B. Dataset

The Improved GMM was implemented using Python, leveraging libraries such as NumPy for numerical computations, OpenCV for video processing and background subtraction, Scikit-learn for Gaussian Mixture Model operations, and Matplotlib for result visualization. Jupyter Notebook served as the primary development environment, enabling interactive code execution and analysis. The experiments were conducted on a standard laptop equipped with an Intel Core i5 processor (2.50 GHz), 8 GB RAM, and Windows 11 (64-bit). The system achieved a processing speed of approximately 0.75 seconds per frame.

The evaluation of the Improved GMM was conducted using the CDNet 2014 dataset, a benchmark dataset widely used for background subtraction methods. This dataset includes videos representing various traffic densities, categorized as light traffic (less than three vehicles per 500 square feet), moderate traffic (three to five vehicles per 500 square feet), and heavy traffic (more than five vehicles per 500 square feet).

To prepare the dataset for analysis, each video was divided into individual frames for frame-by-frame processing. The

frames were resized to a consistent resolution of 640×480 pixels to ensure efficient processing. Ground truth masks provided in the dataset were used for validation by comparing them against binary masks generated by the Improved GMM. The implementation of the Improved GMM algorithm followed the modifications outlined in Section III. Key parameters were initialized. The CS algorithm was employed to automatically tune these parameters, ensuring optimal performance for various traffic conditions. During frame processing, the GMM dynamically updated its parameters using adaptive learning rates and exponential decay to refine the model, while outlier processing mitigated noise and sudden changes. Finally, the binary masks were refined through morphological operations, such as dilation and erosion, to eliminate noise and enhance detection accuracy. The experiment involved the Improved GMM method and was compared with masks generated using the traditional GMM [6], Effective Adaptive GMM (EGMM) [14], Region-based Mixture of Gaussians (RMoG) [15], Boosted GMM (BMOG) [25], Competitive Learning for Varying Input Distributions (CL-VID) [45], Real-Time Sematic Background Subtraction Version 2 (RT-SBS-V2) [16] and Common Vector Approach Background Subtraction (CVABS) [17].

C. Cuckoo Search Optimization

The CS optimization algorithm was employed to automatically tune the key parameters of the Improved GMM, including $c, K, \alpha, \lambda, \gamma$, and threshold (th). This algorithm, inspired by the breeding behavior of cuckoos, efficiently explores the parameter space using a balance between local and global search mechanisms. To determine the appropriate parameter ranges, prior literature and empirical experimentation were consulted. Table I provides the initial ranges for each parameter:

TABLE I. RANGE PARAMETERS OF OPTIMIZED PARAMETERS

Parameters	Range
c	$0.01 \leq c \leq 0.1$
K	$2.0 \leq K \leq 5.0$
α	$5.0 \leq \alpha \leq 50.0$
λ	$0.01 \leq \lambda \leq 1.0$
γ	$0.1 \leq \gamma \leq 1.0$
th	$3.0 \leq th \leq 7.0$

The CS algorithm began by initializing a population of candidate solutions, each representing a unique set of parameter values within the predefined ranges. These candidates were evaluated using a fitness function designed to maximize the accuracy of the Improved GMM's background subtraction. The fitness function compared the binary masks generated by the model with the ground truth annotations, considering metrics such as precision, recall, and F-measure. During each iteration, the algorithm updated the candidate solutions by simulating the levy flights of cuckoos, which allow for both local fine-tuning and global exploration of the parameter space. Poor-performing solutions were replaced by better-performing ones, ensuring that the algorithm converged towards an optimal set of parameter values.

Default settings for CS algorithm hyperparameters were replaced with experimental investigations to determine the most effective parameter combinations for the Improved GMM. Specifically, the number of Gaussian components (N) and the probability of adaptation (Pa) were key factors in optimizing the background modeling process.

The number of Gaussian components (N) represents the number of distributions used to model each pixel's background, which is crucial in capturing the complexity of background variations. A higher N allows for more sophisticated background modeling, enabling the algorithm to better handle complex scenes with multiple background states such as changing lighting conditions, moving shadows, or repetitive background patterns. The probability of adaptation (Pa) controls the rate at which the model updates its background distributions. By experimenting with different values of N and Pa, we aim to find the optimal configuration that provides the most accurate and robust background subtraction across various traffic scenarios. Table II and Table III illustrate the experimental results for different values of N and Pa.

TABLE II. EXPERIMENTAL RESULTS FOR DIFFERENT VALUES OF N USING PA=0.25 OVER CDNET2014 DATASET

Frame	Input	N=10	N=30	N=50
#0672				
#0808				
#1328				
#1517				

TABLE III. EXPERIMENTAL RESULTS FOR DIFFERENT VALUES OF PA USING N=30 OVER CDNET2014 DATASET

Frame	Input	Pa=0.25	Pa=0.5	Pa=0.75
#0672				
#0808				
#1328				
#1517				

Table II presents the experimental results for different values of N (number of Gaussian components in the GMM) while keeping the probability parameter Pa = 0.25 constant. It showcases how varying the number of Gaussian components

affects the background subtraction performance across multiple frames. Increasing N typically balances model complexity and overfitting, influencing detection accuracy. Table III evaluates the impact of varying probability thresholds for background classification using a fixed N=30. The results demonstrate how different threshold values influence the system's sensitivity to classify pixels as foreground or background, emphasizing the importance of parameter tuning for effective detection in varying conditions.

The combined insights from these tables underscore the sensitivity of the GMM model to these parameters and highlight the necessity for optimization methods, such as CS Optimization, to automatically determine the best parameter values for achieving optimal detection performance. Through the CS optimization method, the algorithm converged on an optimal set of parameters, as shown in Table IV:

TABLE IV. OPTIMIZED PARAMETER OF IMPROVED GMM AUTOMATIC TUNING BY CS OPTIMIZATION METHOD

c	K	α	λ	γ	th
0.1	2.0	9.1	1.0	0.6	3.2

The effectiveness of the automatic parameter tuning was compared against empirical tuning, as demonstrated in Table V. The automatic tuning approach demonstrates improved adaptability, producing masks closer to the ground truth across different frames. The empirical tuning approach may suffer from inconsistency due to the lack of parameter optimization for specific scenarios. The visual comparison indicates that automatic tuning effectively reduces background noise and captures more accurate object contours, particularly in challenging scenarios.

TABLE V. COMPARISON OF FOREGROUND MASKS OF AUTOMATIC TUNING PARAMETER OF IMPROVED GMM AND EMPIRICAL TUNING PARAMETER OF IMPROVED GMM

Frame	Input	Ground Truth	Automatic Tuning	Empirical Tuning
#0672				
#0808				
#1328				
#1517				

Table VI evaluates the average performance metrics of the Improved GMM with automatic tuning (optimized parameters) and empirical tuning (manually set parameters). The values demonstrate that the automatic tuning of parameters using CS optimization substantially outperforms empirical tuning. The improved precision reduced false positives, and higher overall accuracy highlight the importance of optimization in achieving robust and reliable vehicle detection.

TABLE VI. AVERAGE PERFORMANCE PARAMETER EVALUATION OF AUTOMATIC TUNING IMPROVED GMM AND EMPIRICAL TUNING IMPROVED GMM

Tuning	RC	PR	FM	FPR	FNR	WC	AC
Automatic	0.533	0.634	0.579	0.033	0.467	0.077	92.3%
Empirical	0.480	0.292	0.422	0.232	0.160	0.225	77.5%

Fig. 1 provides a visual comparison of the accuracy values between automatic and empirical tuning methods. The accuracy achieved with automatic tuning is 92.3%, as observed in Table VI. This high value underscores the effectiveness of automatic tuning in optimizing the Improved GMM parameters for robust and consistent detection across frames. Empirical tuning achieves an accuracy of 77.52%, significantly lower than automatic tuning. The drop in performance highlights the limitations of manually setting parameters, which are less adaptable to variations in input conditions such as different lighting, occlusions, or traffic densities. The results validate the incorporation of CS optimization as a robust method for parameter optimization in vehicle detection tasks.

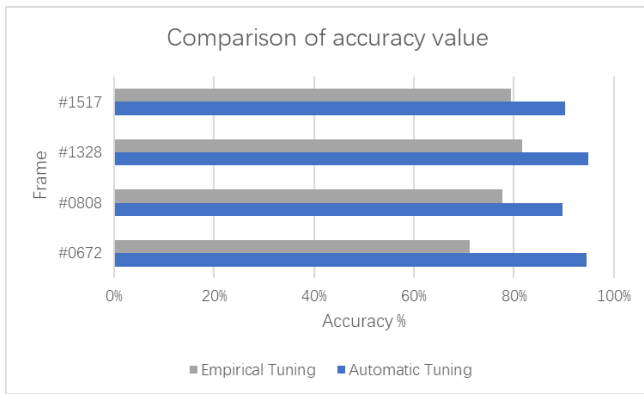


Fig. 1. The core principle of the BS method (a) input image (b) reference image (c) foreground or background.

D. Evaluation Metrics

The poor performance of methods that use GMM for background subtraction is mostly caused by assumptions about the parameters, where different issues produce varying effects on methods performance. Additionally, performance evaluation metrics showcase the advantages of background subtraction methods in vehicle detection accuracy. The pixel features from vehicle detection results were divided into two groups (foreground considered positive and background considered negative) to evaluate the detection accuracy objectively [18].

Pixel-based performance evaluation metrics were widely employed to assess the accuracy of image segmentation algorithms. These metrics compared the pixel-wise segmentation results obtained by the algorithm to the ground truth (GT) annotations. Once the GT was determined, several generally accepted methods were compared to a proposed binary foreground map. In this evaluation, four types of pixels are utilized as follows:

- True positive (TP)
- False positive (FP)

- False negative (FN)
- True negative (TN)

The TP represents pixels correctly classified as positive by the segmentation algorithm and positive in the GT annotation. Meanwhile, FP represents pixels classified as positive by the segmentation algorithm but are negative in the GT annotation. The FN represents negative pixels in the segmentation algorithm but positive in the GT annotation. Subsequently, TN represents pixels correctly classified as negative by the segmentation algorithm and negative in the GT annotation [46].

Based on these four types of pixels, several metrics were computed to evaluate the performance of the segmentation algorithm, including recall (RC), precision (PR), F-measure (FM), FP rate (FPR), FN rate (FNR), and accuracy (AC). The RC is a metric used to evaluate the ability of an algorithm to correctly identify positive pixels, which is calculated as the ratio of TP pixels to the total number of ground TP pixels (also known as a TP rate). The TP rate measures the fraction of foreground pixels accurately identified out of the total number of foreground pixels, which the algorithm has categorized as [47]:

$$RC = TP / (FN + TP) \quad (19)$$

Since PR is the fraction of TP pixels over the number of positive pixels classified by the segmentation algorithm, the proportion of positive predictions that are TPs is measured given by:

$$PR = TP / (TP + FP) \quad (20)$$

The FM, or F1 score, is the harmonic mean of PR and RC. Therefore, FM provides a single metric to evaluate the overall performance of the algorithm as follows:

$$FM = (2 \times RC \times PR) / (RC + PR) \quad (21)$$

The FPR is the fraction of FP pixels over the total GT negative pixels. Hence, the proportion of negative predictions that are FPs is expressed as:

$$FPR = FP / (FP + TN) \quad (22)$$

In FNR, the fraction of FN pixels over the total number of GT positive pixels is obtained. Therefore, the proportion of positive pixels that are missed by the algorithm is described by:

$$FNR = FN / (FN + TP) \quad (23)$$

The AC is the fraction of correctly classified pixels over the total number of pixels and measures the overall correctness of the segmentation of the algorithm is represented by:

$$AC = (TP + TN) / (TP + FP + FN + TN) \quad (24)$$

Finally, wrong classification (WC) is the fraction of wrongly classified pixels over the total number of pixels. The WC also measures the overall error rate of the segmentation of the algorithm is expressed by:

$$WC = (FP + FN) / (TP + FP + FN + TN) \quad (25)$$

These seven indicators explain and evaluate the performance of the proposed model. Generally, these metrics provide a quantitative evaluation of the performance of the segmentation algorithm while aiding in identifying areas where the algorithm requires improvement [48].

VI. RESULT AND DISCUSSION

To validate the effectiveness of the proposed Improved Gaussian Mixture Model (GMM) [29], extensive experiments were conducted using over 1700 frames of moving vehicles on the road. Table VII compares the Improved GMM method with several state-of-the-art background subtraction and vehicle detection methods. The table includes the evaluation metrics used in this study and the results obtained for each method under different traffic densities.

TABLE VII. SUMMARY OF GMM, EGMM, RMOG, BMOG, CL-VID, RT-SBS-V2, CVABS AND IMPROVED GMM AVERAGE METRICS

Method	RC	PR	FM	FPR	FNR	AC	WC
GMM [6]	0.972	0.806	0.891	0.026	0.002	0.970	0.024
EGMM [14]	0.970	0.783	0.876	0.029	0.004	0.970	0.027
RmoG [15]	0.970	0.662	0.795	0.047	0.005	0.957	0.043
BMOG [25]	0.949	0.814	0.906	0.026	0.002	0.974	0.023
CL-VID [45]	0.967	0.849	0.922	0.019	0.013	0.975	0.018
RT-SBS-V2 [16]	0.973	0.790	0.881	0.031	0.004	0.969	0.028
CVABS [17]	0.916	0.849	0.886	0.020	0.084	0.972	0.028
Improved GMM	0.959	0.835	0.890	0.026	0.041	0.973	0.027

A high RC value was desirable to minimize FNs. The FNs occurred when actual positive instances were incorrectly classified as negative, presenting missed detections or opportunities. Thus, a high RC indicated that the model effectively captured the majority of positive instances and was less likely to overlook essential or critical cases [48]. The RC value of the Improved GMM outperformed a few methods with a value of 0.9586. This improvement is particularly impactful in scenarios with heavy traffic densities, where high vehicle overlap increases the likelihood of misclassification in traditional methods. The adaptability of the Improved GMM allows it to differentiate subtle variations in pixel distributions, reducing false negatives in densely packed vehicle scenarios, such as urban intersections or peak highway traffic conditions.

Alternatively, a high PR value suggested that the methods produced a lower FP rate, which could accurately identify positive pixels with a lower tendency to incorrectly classify background pixels as foreground. In other words, a high PR denoted that the methods of a pixel as foreground was more likely to be accurate. The PR value for the Improved GMM was 0.8353, which only slightly differed from the highest values from CL-VID and CVABS. This demonstrates the Improved GMM's effectiveness in reducing false positives, particularly in moderate traffic conditions where objects such as pedestrians, shadows, or reflections might otherwise be misclassified. By maintaining a robust precision level, the Improved GMM ensures reliable vehicle detection, crucial for real-world applications like urban traffic monitoring or congestion management.

The FM is a widely used metric that combines PR and RC to evaluate the overall performance of a method. A higher FM indicated that the methods significantly balanced PR and RC, effectively identifying positive pixels while minimizing FPs. Therefore, a higher FM value suggested better performance. Meanwhile, the FM value for the Improved GMM was among the highest among the other state-of-the-art methods. This positions the Improved GMM as a balanced performer, particularly in traffic scenarios with moderate density, where both precision and recall are crucial for maintaining detection reliability. Compared to CL-VID and BMOG, which excel in heavy and light traffic respectively, the Improved GMM demonstrates consistent and reliable performance across varied traffic densities, making it a versatile choice for real-world applications. The optimal FM value depended on the specific application and the desired trade-off between PR and RC [48].

A low FNR was considered desirable in most situations in which the system accurately identified positive instances and minimized missed detections or incorrect negative predictions. For example, the Improved GMM achieved a lower FNR compared to traditional GMM (0.041 vs. 0.084), highlighting its superior ability to identify true positives effectively. This improvement is particularly beneficial in high-density traffic scenarios, where the risk of missed vehicle detections is higher due to occlusions. Like FPR, the desired FNR depended on the specific context and application [49]. In some cases, a trade-off could occur between the FNR and other factors, including FPR or the overall system cost. The appropriate balance could vary depending on the situation's specific goals, constraints, and acceptable risks. Nonetheless, a low FNR was generally preferable to ensure the highest AC and detection performance [49].

The AC metric measures the percentage of correctly classified pixels in the foreground mask. A high AC suggested that the methods successfully classified a larger proportion of positive and negative pixels relative to the total number of pixels. This value accurately reflected the ability of the methods to distinguish between foreground and background regions. Lastly, WC referred to incidents in which a classification or prediction system improperly categorized them. For instance, in a real-world traffic monitoring system deployed at a busy urban intersection, high WC rates could lead to incorrect vehicle counts or misclassifications, potentially compromising traffic flow optimization or accident detection. The Improved GMM's ability to minimize WC ensures more accurate vehicle detection and tracking, leading to improved reliability in such critical applications. Depending on the applications, high WC rates could produce errors, misinterpretations, and negative consequences [48].

Fig. 2 illustrates the comparative recall performance of eight computational methods across three different traffic densities: light, moderate, and heavy. Recall is a metric that measures the percentage of true positive detections from all the actual positive instances in the video. Recall measures the percentage of true positive detections from all actual positive instances in the video. The graph reveals that some methods, such as CL-VID and RT-SBS-V2, maintain consistently high recall in specific conditions, while others show greater variability. For instance, GMM demonstrates a slight edge in

heavy traffic scenarios, likely due to its strong adaptability to high-density environments. Meanwhile, RT-SBS-V2 excels in light conditions, benefiting from its sensitivity to minimal background dynamics. Conversely, the CVABS method demonstrates a notable decline in recall under moderate traffic conditions compared to its performance in light and heavy conditions. This inconsistency may indicate specific weaknesses, such as sensitivity to moderately complex environments with fluctuating vehicle densities or partial occlusions. By analyzing the recall trends across traffic densities, Fig. 2 highlights the strengths and areas for improvement in the Improved GMM, emphasizing its potential as a balanced performer for diverse real-world applications.

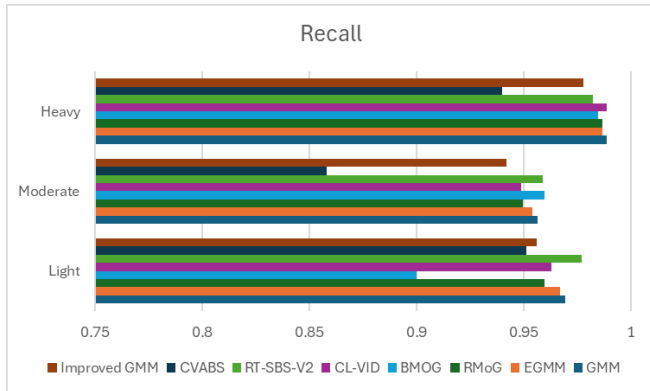


Fig. 2. The comparison of recall results of our proposed method and the other methods.

The PR values are a key performance indicator, particularly in fields with high costs of false positives. Fig. 3 shows that the method labelled CL-VID shows the highest PR value in the heavy category, underscoring its exceptional accuracy under challenging conditions where vehicle overlap, and background complexity are prominent. This suggests that CL-VID is particularly adept at minimizing false positives when the scene dynamics are most intricate. In contrast, RMoG exhibits lower PR values, especially in the moderate traffic category. This performance discrepancy could be attributed to its limitations in handling medium-complexity scenarios, such as moderate occlusions or partially visible vehicles. The proposed Improved GMM achieves consistently high precision across all traffic densities, highlighting its robustness and versatility. While it does not outperform CL-VID in the heavy category, its balanced performance across light, moderate, and heavy traffic conditions signifies its reliability as a general-purpose solution for diverse environments. This generalist approach ensures that the Improved GMM maintains low false-positive rates regardless of traffic complexity, making it a dependable choice for real-world applications.

BMOG and CVABS methods show competitive PR values, particularly in moderate and heavy traffic scenarios. Their ability to sustain high precision in demanding conditions emphasizes their effectiveness in environments with increased vehicle density and dynamic lighting changes. A noteworthy observation is that PR values exhibit less variation across methods compared to RC. This stability indicates that these methods generally maintain a consistent ability to identify true positives. However, differences in precision suggest that their

effectiveness in rejecting false positives varies, which is critical in applications requiring high reliability and minimal false alarms.

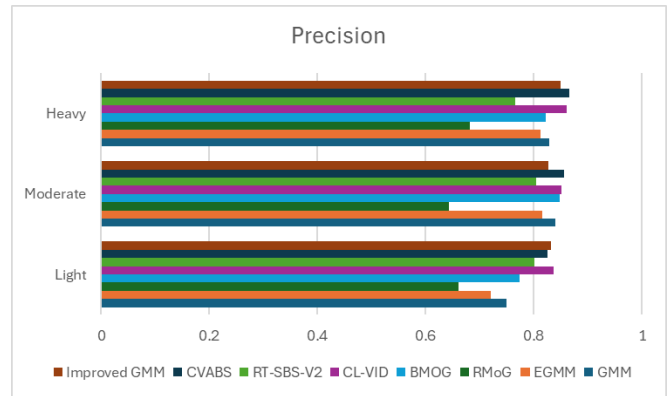


Fig. 3. The comparison of precision results of our proposed method and the other methods.

Fig. 4 illustrates the FM results of eight different background subtraction methods: GMM, EGMM, RMoG, BMOG, CL-VID, RT-SBS-V2, CVABS, and the proposed method. As a harmonic mean of precision and recall, FM provides a comprehensive measure of segmentation accuracy, making it a crucial metric for evaluating the effectiveness of these methods. The graph highlights the superior performance of CL-VID in scenarios with moderate and heavy background motion, showcasing its robustness and adaptability in handling complex conditions such as dynamic lighting and high traffic density. This positions CL-VID as a strong candidate for scenarios requiring high reliability in challenging environments. In simpler conditions, the BMOG method achieves the highest FM in the light traffic category. This indicates its ability to excel in straightforward segmentation tasks where background dynamics are less pronounced, making it suitable for environments with minimal motion complexity.

The Improved GMM demonstrates consistent performance across all traffic categories. While it does not achieve the top FM score in any specific category, its stability across light, moderate, and heavy conditions highlights its versatility and reliability as a general-purpose solution. This balanced performance makes the Improved GMM particularly suitable for applications requiring robust results across diverse operational scenarios. In contrast, RMoG exhibits lower FM values across all categories, signalling potential limitations in adapting to varying background complexities. This underperformance may stem from its inability to handle dynamic textures or abrupt changes effectively. A general trend observed from the results is that most methods perform better in light and moderate categories, with a slight decline in heavy traffic conditions. This trend underscores the increasing challenge posed by higher background complexity and overlapping objects in heavy traffic scenarios, which can impact segmentation accuracy.

The stable performance of the Improved GMM across different conditions underscores its potential for applications requiring consistent and reliable segmentation, such as real-time traffic monitoring and video analytics. While other

methods, such as CL-VID and BMOG, excel in specific conditions, the Improved GMM offers a balanced approach, ensuring dependable performance irrespective of environmental variability.

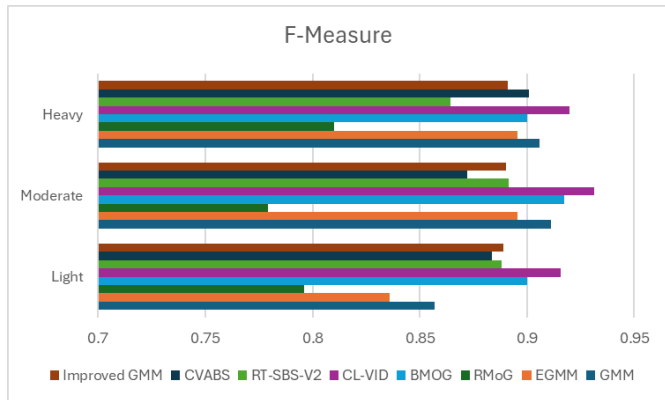


Fig. 4. The comparison of F-measure results of our proposed method and the other methods.

Fig. 5 illustrates the AC results of various background subtraction methods under different traffic conditions. The data reveals that all methods achieve high accuracy, with most exceeding 0.95, reflecting their overall effectiveness in segmentation tasks. However, noticeable differences emerge when comparing performances across specific traffic conditions. BMOG and CL-VID stand out with superior accuracy in light and moderate conditions, respectively, demonstrating their specialised efficiency in less complex environments. Their performances converge under heavy traffic conditions, suggesting that their methodologies are similarly adept at handling challenging scenarios with high background complexity and vehicle overlap. The proposed Improved GMM maintains consistently high accuracy across all traffic densities, highlighting its robustness and adaptability. This consistency suggests that the Improved GMM effectively balances precision and recall, making it well-suited for diverse operational contexts, including scenarios with fluctuating background dynamics. While the RMoG method achieves relatively high accuracy, its lower performance in moderate and heavy conditions indicates potential limitations in handling dynamic or complex backgrounds. This shortfall could be attributed to its reduced ability to effectively manage abrupt changes or intricate textures.

The subtle variations in accuracy among these methods carry significant implications for practical applications where precision is critical. For example, in high-security environments or scenarios requiring detailed video analysis, even small accuracy differences can influence the choice of method. The slight decline in accuracy under heavy traffic conditions observed for most methods emphasizes the need for enhanced robustness against complex backgrounds, direction future research should explore. The consistent performance of the proposed Improved GMM suggests a well-balanced integration of techniques tailored to address varying complexities. This adaptability makes it a compelling option for researchers aiming to develop versatile background subtraction methods that perform reliably across diverse conditions.

Table VIII compares foreground masks obtained using eight different methods. The table shows each method's original frame (input), ground truth (GT), and segmentation masks. The frames displayed are from three different traffic densities:

- 1) Frame 1213: Heavy traffic scenario
- 2) Frame 1480: Moderate traffic scenario
- 3) Frame 795: Light traffic scenario

The first column displays the original input frames, representing the raw video data captured by the camera. The second column shows the GT, which consists of manually annotated masks indicating the precise locations of vehicles in the respective frames. The subsequent columns (third to tenth) present segmentation masks produced by the different methods, allowing a direct comparison against the GT and input frames.

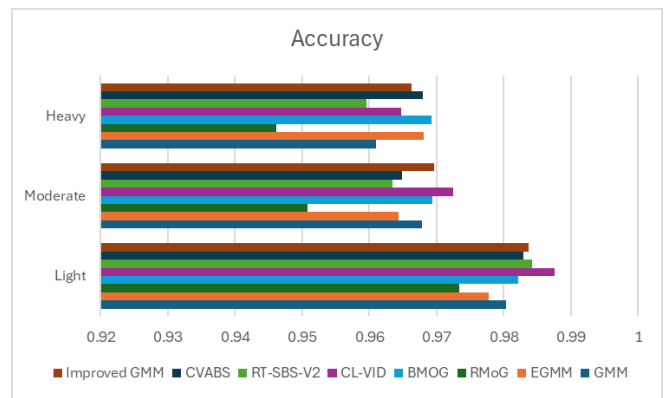
































Fig. 5. The comparison of accuracy results of our proposed method and the other methods.

As illustrated in Table VIII, the Improved GMM demonstrates satisfactory detection performance across all traffic densities. For example, in the heavy traffic scenario, the Improved GMM effectively captures vehicle locations and shapes with notable accuracy, achieving a balance between minimizing false positives and negatives. While the Improved GMM exhibits competitive performance, the results also reveal the distinct strengths and weaknesses of other methods. For instance, CL-VID and BMOG show strong performance in scenarios with moderate traffic, excelling in precision and segmentation clarity. Meanwhile, EGMM and CVABS perform well in light traffic scenarios, accurately identifying individual vehicles with minimal background noise.

False positives and false negatives are present to varying degrees across all methods, including the Improved GMM. This indicates the inherent challenges of background subtraction in dynamic traffic environments, such as occlusions, varying lighting conditions, and overlapping vehicles. The Improved GMM method aims to strike a balance between computational efficiency and segmentation accuracy, making it a viable option for real-time applications where processing time is a critical constraint. However, the increased complexity introduced by the Improved GMM, particularly due to features like adaptive learning rates and exponential decay, warrants further investigation to fully assess its scalability and efficiency under diverse operational scenarios.

TABLE VIII. COMPARISON OF SEGMENTATION MASKS FOR THE CDNET2014 DATASET

Traffic	Light Traffic	Moderate Traffic	Heavy Traffic
Input			
GT			
GMM [6]			
EGMM [14]			
RMoG [15]			
BMOG [25]			
CL-VID [45]			
RT-SBS-V2 [16]			
CVABS [17]			
Improved GMM			

VII. CONCLUSION AND FUTURE WORK

This study proposed an Improved GMM for high-accuracy vehicle detection across varying traffic densities. By incorporating an adaptive time-varying learning rate, exponential decay, and outlier processing, the model effectively addressed limitations in traditional GMM methods, such as fixed learning rates and sensitivity to outliers. The use of CS Optimization for automatic parameter tuning further enhanced the robustness and adaptability of the model. Experimental results demonstrated that the proposed method consistently achieved superior performance metrics, including accuracy, precision, recall, and F-measure, across light, moderate, and heavy traffic scenarios.

Despite these promising results, several limitations warrant consideration. First, the computational complexity of the proposed model may present challenges for deployment on low-resource systems. Second, the model's reliance on hand-crafted features might limit its adaptability to rapidly evolving traffic scenarios. Future work should explore integrating deep learning techniques to complement the Improved GMM for enhanced robustness and scalability. Additionally, incorporating real-time optimization and reducing

computational overhead will be critical for broader adoption in resource-constrained environments.

In conclusion, the proposed method offers a significant step toward reliable and efficient vehicle detection in diverse traffic scenarios, paving the way for further advancements in intelligent transportation systems.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and editors for their valuable suggestions and comments. This research was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2021/STG06/UTHM/02/1).

REFERENCES

- [1] D. Y. Ge, X. F. Yao, W. J. Xiang, and Y. P. Chen, "Vehicle detection and tracking based on video image processing in intelligent transportation system," *Neural Computing and Applications*, vol. 35, no. 3, pp. 2197–2209, Jan. 2023.
- [2] K. Zhong, Z. Zhang, and Z. Zhao, "Vehicle detection and tracking based on GMM and enhanced camshift algorithm," *Journal of Electrical and Electronic Engineering*, vol. 6, no. 2, pp. 40–45, Apr. 2018.
- [3] N. A. Mohd Aris and S. S. Jamaian, "Background subtraction challenges in motion detection using Gaussian mixture model: a survey," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 3, pp. 1007-1018, 2023.
- [4] R. Deng, D. Yang, X. Liu, and S. Liu., "A background subtraction algorithm based on pixel state," in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 251–254, Nov. 2014.
- [5] T. Indu, Y. Shivani, A. Reddy, and S. Pradeep, "Real-time classification and counting of vehicles from CCTV videos for traffic surveillance applications," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 14, no. 2, pp. 684–695, May 2023.
- [6] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 246–252, Jul. 2006.
- [7] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. 28–31, Aug. 2004.
- [8] S. Rajkumar, A. Hariharan, S. Girish, and M. Arulmurugan, "An efficient vehicle detection and shadow removal using Gaussian mixture models with blob analysis for machine vision application," *SN Computer Science*, vol. 4, no. 5, pp. 451, Jun. 2023.
- [9] L. Alandkar and S. R. Gengaje, "Dealing background issues in object detection using GMM: a survey," *International Journal of Computer Applications*, vol. 150, no. 5, pp. 0975–8887, Sept. 2016.
- [10] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using a mixture of Gaussians for foreground detection—a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, Nov. 2008.
- [11] N. Friedman and S. Russell, "Image segmentation in video sequences: a probabilistic approach," *13th Conference on Uncertainty in Artificial Intelligence*, 175-181, 1997.
- [12] J. Zuo, Z. Jia, J. Yang, and N. K. Kasabov, "Moving target detection based on improved Gaussian mixture background subtraction in video images," *IEEE Access*, vol. 7, pp. 152612-152623, 2019.
- [13] L. L. Lin, and N. R. Chen, "Moving objects detection based on Gaussian mixture model and saliency map," *Applied Mechanics and Materials*, vol. 63-64, pp. 350-354, 2011.
- [14] Z. Zivkovic and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773-780, 2006.

- [15] S. Varadarajan, P. Miller and H. Zhou, "Spatial mixture of Gaussians for dynamic background modeling," In Proceedings of the 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 63-68, 2013.
- [16] A. Cioppa, M. V. Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Oct. 2020, pp. 3214–3218.
- [17] Ş. Işık, K. Özkan, and Ö. N. Gerek, "CVABS: moving object segmentation with a common vector approach for videos," IET Computer Vision, vol. 13, no. 8, pp. 719–729, Dec. 2019.
- [18] Y. Zhang, C. Zhao, J. He, and A. Chen, "Vehicles detection in complex urban traffic scenes using a nonparametric approach with confidence measurement," in 2015 International Conference and Workshop on Computing and Communication (IEMCON), pp. 1–7, Oct. 2017.
- [19] K..A. B. Lima, K. R. T. Aires, and F. W. P. D. Reis, "Adaptive method for segmentation of vehicles through local threshold in the Gaussian mixture model," Brazilian Conference on Intelligent Systems (BRACIS), pp. 204-209. 2015.
- [20] S. Agrawal and P. Natu, "An improved gaussian mixture method-based background subtraction model for moving object detection in outdoor scene," In Journal of Electrical and Electronic Engineering, vol. 6, no. 2, pp. 40-45, 2021.
- [21] X. Luo, Y. Wang, B. Cai, and Z. Li, "Moving Object Detection in Traffic Surveillance Video: New MOD-AT Method Based on Adaptive Threshold. ISPRS International Journal of Geo-Information, vol. 10, no. 11, pp. 742, 2021.
- [22] Z. Chen and T. Ellis, "Self-adaptive Gaussian mixture model for urban traffic monitoring system," In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1769–1776, 2011.
- [23] K. Kalti and M. A. Mahjoub, "Image segmentation by Gaussian mixture models and modified FCM algorithm," Int. Arab J. Inf. Technol., vol. 11, no. 1, pp. 11-18, 2014.
- [24] H. Wei and W. Zheng, "Image Denoising Based on Improved Gaussian Mixture Model," Scientific Programming, 2021.
- [25] I. Martins, P. Carvalho, L. Corte-Real, and J. L. Alba-Castro, "BMOG: boosted Gaussian mixture model with controlled complexity for background subtraction," Pattern Analysis and Applications, vol. 21, no. 3, pp. 641–654, Aug. 2016.
- [26] Y. Su, "Target detection algorithm and data model optimization based on an improved Gaussian mixture model," Microprocessors and Microsystems, vol.81, pp. 103797, 2021.
- [27] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01), pp. 133–144, Sept. 2002.
- [28] D. S. Lee, "Effective Gaussian mixture learning for video background subtraction," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp. 827–832, Mar. 2005.
- [29] H. Wang and P. Miller., "Regularized online mixture of Gaussians for background subtraction," presented at the 2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 249-254, Aug. 2011.
- [30] J. Fried, F. Lizarralde, and A. C. Leite., "Adaptive Image-based Visual Servoing with Time-varying Learning Rates for Uncertain Robot Manipulators," presented at the 2022 American Control Conference (ACC), pp. 3838-3843, Jun. 2022.
- [31] P. Toulis, T. Horel, and E. M. Airoldi, "The proximal Robbins-Monro method," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 83, no. 11, pp. 188–212, Feb. 2021.
- [32] H. Robbins and S. Monro, "A stochastic approximation method," The Annals of Mathematical Statistics, vol. 22, no. 3, pp. 400–407, Sept. 1951.
- [33] N. A. Mohd Aris, S. S. Jamaian, and D. R. Sulistyningtum, "Vehicle Detection Based on Improved Gaussian Mixture Model for Different Weather Conditions," Journal of Advanced Research in Applied Sciences and Engineering Technology, pp. 160-170, 2024.
- [34] C. W. Liu, B. Andersson, and A. Skrondal, "A constrained Metropolis-Hastings Robbins-Monro algorithm for Q matrix estimation in DINA models," Psychometrika, vol. 85, no. 2, pp. 322–357, Jun. 2020.
- [35] H. A. Tehrani, A. Bakhshi, and T. T. Y. Yang, "Online jointly estimation of hysteretic structures using the combination of central difference Kalman filter and Robbins-Monro technique," Journal of Vibration and Control, vol. 27, no. 1-2, pp. 234–247, Jan. 2021.
- [36] C. M. Bishop, Pattern Recognition and Machine Learning. New York, USA: Springer Google Schola, 2006. Accessed: August 23, 2016. [Online]. Available: <https://link.springer.com/in/book/9780387310732>.
- [37] S. Ruder, An overview of gradient descent optimization algorithms. Accessed: January 19, 2016. [Online]. Available: arXiv preprint arXiv:1609.04747.
- [38] E. Aboutanios, "Estimation of the frequency and decay factor of a decaying exponential in noise," IEEE Transactions on Signal Processing, vol. 58, no. 2, pp. 501–509, Sept. 2009.
- [39] K. Deng, "Exponential decay of solutions of semilinear parabolic equations with nonlocal initial conditions," Journal of Mathematical Analysis and Applications, vol. 179, no. 2, pp. 630–637, Nov. 1993.
- [40] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72–83, Jan. 1995.
- [41] C. V. Stewart, "Robust parameter estimation in computer vision," SIAM Review, vol. 41, no. 3, pp. 513–537, 2005.
- [42] J. Warwick and M. C. Jones, "Choosing a robustness tuning parameter," Journal of Statistical Computation and Simulation, vol. 75, no. 7, pp. 581–588, Jul. 2005.
- [43] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," IEEE Signal Processing Magazine, vol. 29, no. 4, pp. 61–80, Jun. 2012.
- [44] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 1, pp. 439–445, Jun. 2000.
- [45] E. López-Rubio, M. A. Molina-Cabello, R. M. Luque-Baena, and E. Domínguez, "Foreground detection by competitive learning for varying input distributions," International Journal of Neural Systems, vol. 28, no. 5, pp. 1750056, Jun. 2018.
- [46] T. Schlogl, C. Beleznai, M. Winter, and H. Bischof, "Performance evaluation metrics for motion detection and tracking," in Proceedings of the 17th International Conference on Pattern Recognition, pp. 519–522, 2004.
- [47] K. Oksuz, B. C. Cam, E. Akbas and S. Kalkan, "Localization recall precision (LRP): A new performance metric for object detection," in Proceedings of the European Conference on Computer Vision (ECCV), pp. 504–519, 2018.
- [48] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "ChangeDetection.net: A new change detection benchmark dataset," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 1–8, Jun. 2012.
- [49] N. Lazarevic-McManus, J. P. Renno, and G. A. Jones, "Performance evaluation in visual surveillance using the F-measure," in Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks, pp. 45–52, Oct. 2006.