IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 3, September 2023, pp. 1007∼1018
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i3.pp1007-1018

Background subtraction challenges in motion detection using Gaussian mixture model: a survey

Nor Afiqah Mohd Aris, Siti Suhana Jamaian
Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, Malaysia

Article Info

Article history:

Received Jun 25, 2022
Revised Aug 24, 2022
Accepted Nov 2, 2022

Keywords:

Background-foreground detection
Background subtraction
Computer vision
Gaussian mixture model
Motion detection

ABSTRACT

Motion detection is becoming prominent for computer vision applications. The background subtraction method that uses the Gaussian mixture model (GMM) is utilized frequently in camera or video settings. However, there is still more work that needs to be done to develop a reliable, accurate and high-performing technique due to various challenges. The degree of difficulty for this challenge is primarily determined by how the object to be detected is defined. It could be influenced by the changes in the object posture or deformations. In this context, we describe and bring together the most significant challenges faced by the background subtraction techniques based on GMM for dealing with a crucial background situation. Therefore, the findings of this study can be used to identify the most appropriate GMM version based on the crucial background situation.

Corresponding Author:

Siti Suhana Jamaian
Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub
84600 Pagoh, Johor, Malaysia
Email: suhana@uthm.edu.my

1. INTRODUCTION

The study and comprehension of video sequences have become an active area of research in computer vision over the past few years due to their increasing importance in numerous video analysis applications, such as video surveillance and multimedia applications that detect motion in the video scene [1], [2]. Identifying motion in a video sequence is important in target detection [3]–[5] and behaviour interpretation [6], [7]. Thus, the initial operation is distinguishing between the foreground and background objects. It can be done in various ways, depending on the available data and whether the object is moving. Any prior information to detect moving objects from a video sequence is unnecessary. Instead, a series of consecutive frames is required.

Many motion-based video systems still struggle to deal with dynamically changing backgrounds. Dynamic or moving background objects can cause massive false detections [8]. It is a significant contributor to false alarms in event detection. In order to raise the sensitivity to important events of interest, the security industry has a tremendous need to either detect and suppress these false alarms or mitigate the effects of background changes.

Many new approaches for identifying motion have received attention. In the process of detecting motion, the optical flow method [9]–[12], the inter-frame difference method [13]–[16], and the background subtraction method [17]–[20] are the most used approaches. Detecting methods are chosen based on detection scenarios. The inter-frame difference method extracts a moving target from a continuous video frame image.

It is pretty adaptive. However, this approach causes a cavity effect, reducing target detection accuracy. The optical flow approach uses the target image's brightness. This technology is rarely employed because of its processing complexity and lack of anti-jamming capacity. Meanwhile, background subtraction focuses on building a stable background model to detect motion.

Subtracting the background from a video sequence is one of the most straightforward approaches to finding the motion inside the video. It is an essential procedure for the majority of computer vision applications. In its most basic form, a background subtraction method requires a consistent background, which is an extremely challenging requirement for applications that run in real time. The video sequence is separated into its component video frames, and then each video frame is removed from a background model or reference image. It is assumed that the moving object is represented by the current active structure's pixels distinct from the background model. The foreground object goes through additional processing for object localization and tracking. Because removing the background is the initial step in many applications involving computer vision, it is essential that the obtained foreground pixels precisely correlate to the moving object of interest.

Background subtraction is a method that uses a fairly straightforward algorithm. However, it is extremely unstable to shifts in the surrounding environment and has weak anti-interference capabilities. Many researchers have presented a variety of background subtraction approaches to deal with a variety of problems. Different approaches to simulating the background can be divided into pixel-based, region-based, and hybrid approaches. In addition, there is the possibility of classifying modelling techniques for the background as either parametric or nonparametric. Basic models [21], background estimation [22], background clustering [23], subspace learning, kernel density estimation, and the Gaussian mixture model (GMM) are some video background modelling methodologies that are often utilized in the video. Among these, the GMM technique has gained much interest due to its ability to easily handle image or video noise, shadow, camouflage, slow-moving object, multimodal background and illumination changes, which is particularly noteworthy. However, researchers are still investigating and making new contributions to the established study to increase the object-detection performance. The purpose of this work is to provide a summary of researches that has successfully done on GMM-based object detection in relation to various background environments. The other parts of this paper are structured as follows. In section 2, the GMM is explained and introduced. In section 3, we comprehensively summarise the research on the various GMM versions with respect to various background conditions. Section 4 presents our conclusions.

2. GAUSSIAN MIXTURE MODEL

GMM was first introduced by [24], which is a form of modelling technique based on the background that creates a robust tracking system to handle multiple moving objects, variations in lighting, moving scene clutter, and other arbitrary changes to the scene. The fundamental concept behind this method is to create a Gaussian distribution for every pixel contained within a frames-long series. The foreground and background are each indicated by a different weight, with the foreground having a smaller weight than the background. If the new pixel fits the parameters of the K Gaussian model, then it will be treated as a background pixel. If there is no match for the K Gaussian model, the new pixel will be handled as though it were a foreground pixel. The random variable X is assigned to each pixel and will be modelled as a mixture of K Gaussian distributions, as shown in (1).

$$P(X_t) = \sum_{k=1}^{K} W_{k,t} \cdot \eta\left(X_t, \mu_k, \Sigma_{k,t}\right). \tag{1}$$

$\eta(X_t, \mu_k, \Sigma_{k,t})$ where is the weight of k^{th} Gaussian in the mixture of time t, $W_{k,t}$ is the Gaussian probability density function with mean μ_k and covariance matrix $\Sigma_{k,t}$ which is given by (2).

$$\eta(X_t, \mu_k, \Sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\sigma_{k,t}|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(X_t - \mu_{k,t})^T \sum_{k,t}^{-1}(X_t - \mu_{k,t})\right\} \tag{2}$$

where n is the pixel intensity dimension. It is assumed that the covariance matrix is

$$\Sigma_{k,t} = \sigma_{k,t}^2 I \tag{3}$$

for computational reasons and K is also determined by the computational power and the available memory. Meanwhile, the value of pixels can be modelled by a Gaussian mixture distribution value of K from 3 to 5. These methods assume that the pixel values with red, green, and blue (RGB) have the same variance and are independent. This assumption can avoid a costly matrix inversion at some accuracy rate. As a result, the GMM is applied in this research in order to characterize the arrangement of pixels found in the scene.

The background model estimation process starts when every pixel value is compared to the current Gaussian distribution value K until a match is found. A pixel value within 2.5 standard deviations of the distribution is considered a match. If the match distribution found for the new pixel value is one of the background models, it is regarded as the background. Otherwise, the value of the pixel is the foreground. The B Gaussian distribution, which is chosen as the background model and exceeds a certain threshold which can be written as (4)

$$B = argmin\left(\sum_{b}^{k=1} W_k > Threshold\right). \tag{4}$$

The Gaussian parameters such as the weight, the mean, and the variance must be updated for the subsequent foreground detection. The weight, the mean, and the variance of the k^{th} Gaussian in the mixture at the time t are updated as follows

$$W_{k,t} = (1-\alpha)W_{k,t-1} + \alpha(N_{k,t}) \tag{5}$$

$$\mu_t = (1-\rho)\mu_{t-1} + \rho(X_t) \tag{6}$$

$$\sigma_t^2 = (1-\rho)\mu_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \tag{7}$$

where α is the learning rate and ρ is a second learning rate which is equal to $\alpha\eta(X_t, \mu_k, \sigma_k)$. If the model matched, the value of $N_{k,t}$ is 1 and 0 for the remaining models. If the model is unmatched, the mean and the variance remain unchanged. After making this first approximation, the weights are subsequently renormalized. Suppose the current pixel does not match with the k^{th} Gaussian in the mixture. In that case, the distribution with the most negligible probability is replaced with a distribution that has a current value as its mean value, high initial variance, and low current weight. The value of ω/σ^2 is used to rank the Gaussians. This value will increase as the distribution obtains more evidence and the variance decreases.

3. GMM FOR BACKGROUND SUBTRACTION ISSUES

GMM is able to establish multiple distributions rather than clearly modelling the pixel values as one particular type of distribution [25]. GMM has become a classical parametric model for moving detection and is classified as background and foreground due to its effectiveness and robustness over diverse distribution [26]. It is possible to have a static or dynamic background. It is straightforward to recognize objects against a static background. When there is motion in the background, the object detection system that is needed is quite complicated. The object identification system needs to be able to accommodate dynamic backgrounds while maintaining satisfactory performance in real time. In the next subsections, we will describe various different background scenarios and the solutions for these issues using a GMM-based object detection system.

3.1. Image or video noise

The image size of motion detection can vary due to the effects of camera imaging characteristics, and the noise can easily influence the findings of motion detection. Because of this, the precision of the identification of moving objects is easily compromised [27], [28]. The presence of noise in an image will not only have an impact on how the image appears to the human eye, but it will also have an impact on the following processing of the image, such as the extraction of image features, the categorization, and recognition of images, and so on. Therefore, before processing the image, it is required to do denoising processing on the obtained image. These issues will enhance the image quality and make it easier to process the image once it has been acquired [26].

Chen and Ellis [29] used a multi-dimensional Gaussian kernel density transform (MDGKT) pre-processor to reduce noise in the spectral, temporal, and spatial domains. Each spectral component is smoothed spatially and temporally using a multivariate kernel that can be thought of as the product of two radially symmetric kernels. The Euclidean metric enables a single bandwidth setting for each domain [30], [31],

$$K_{g_a,g_b}(x) = \frac{M}{g_a g_b} k\left(\left\|\frac{X^a}{g_a}\right\|^2\right) k\left(\left\|\frac{X^b}{g_b}\right\|^2\right) \tag{8}$$

where the spatial component, denoted by x^a , and the temporal component marked by x^b , make up the feature vector x. g_a and g_b are the bandwidths of the kernel, and M is the constant of associated normalization. This MDGKT plays a vital role as a pre-processor to enhance the GMM's reliability. It is possible to control the size of the kernel with nothing more than a pair of bandwidth parameters (g_a, g_b), which in turn determines the time interval and the resolution of the GMM. This adjustment significantly affects the background subtraction's efficiency and precision without compromising spatial flexibility.

Kalti and Mahjoub [32] assumes the incorporation of a fuzzy distance into both the expectation-maximization (EM) and adaptive distance-based fuzzy-C-means (ADFCM) algorithms. In this research, the characterization of pixels is based on two features. The first feature explains the inherent attributes of the pixel, and the second feature defines the entire neighbourhoods of the pixel. After that, the classification is determined based on adaptive distance, which gives preference to either one of two attributes depending on where the pixel is located in the picture in terms of its spatial location. The acquired findings have demonstrated that their method performs considerably better than the conventional fuzzy-C-means (FCM) and EM, particularly with regard to the toughness of the face-to noise and the precision of the edges between areas.

Zuo *et al.* [33] improved the conventional GMM for noise interruption by proposing a new method considering the image block averaging method, wavelet semi-threshold, and adaptive background updating method. At the step of background modelling, the video frame is blocked to enable computation and enhance the speed at which modelling is performed. The model of the background is rebuilt by employing the method of image block averaging. The following step involves utilizing wavelet semi-threshold function denoising in conjunction with mathematical morphology closed operation. The noise problem is successfully eliminated, and detection performance is improved in the detection stage of moving targets. During background updating phase, the method of adaptive background updating is used to bring the background update to produce more accurate detection results. The new approach in this work is both subjectively and objectively preferable to the conventional GMM, which validates the efficacy of the system and demonstrates its flexibility.

Wei and Zheng [28] calculate the L2 norm between the GMM corresponding to the two pixels to determine the degree of similarity between the two sets of pixels. The grayscale of the pixel and the abundance of features in the local region of the picture are both represented by the GMM of the image pixels in the image area. Measurements of the pixel grey intensity and the variety of information in the immediate region of the image can be made with more precision if they are based on the difference between individual pixels. Because of the similarities, the performance of the model used to denoise the image is improved, and the image's detailed information is preserved.

Luo *et al.* [27] used a method for detecting motion that considers the variation in the spatial image threshold. The researcher calculates the projected size of motion in the image regions by establishing the mapping relationship between the geometric features of motion in the image regions and the reasonable level circumscribed rectangle (BLOB) of motion in the geographic space. This method is able to set an adaptive threshold for each motion in order to remove unwanted noise during the process of motion detection.

3.2. Sudden illumination change

It is generally accepted that background models can accommodate slow but steady alterations in how the environment appears. For instance, the amount of light in outdoor environments shifts throughout the day. Variations in the scene's illumination can also occur all of a sudden. The sudden turning on or off the light in an interior space is one example of this shift that can take place. These issues can also occur outside scenarios, which as a sudden shift from cloudy to sunlight. The amount of illumination significantly impacts how the background appears and can lead to the detection of false-positive [34].

Chen and Ellis [35] proposed a global illumination of background model adaption and an online dynamical learning rate in order to deal with this challenge. The researchers devised a revolutionary approach using a revised adaptive strategy in the iterative Zivkovic-Heijden GMM (ZHGMM) learning procedure. In this step, they introduce the global illumination change of the median of quotient (MofQ) factor h among the previously learned background and the binary input picture. The MofQ global illuminating altering factor i_c between the current picture and the reference picture i_r is described as follows for all of the pixels that make up a set S,

$$h = median_{s \in S} \left(\frac{i_{c,s}}{i_{r,s}} \right). \tag{9}$$

the mixture model introduces a counter denoted by the letter c for each Gaussian component. The counter c keeps track of how many data sets have been sent to the Gaussian estimating algorithms, while the factor h monitors how the global illumination varies over time. This adjustment makes a significant improvement to the convergence as well as the background subtraction accuracy while maintaining the same degree of temporal flexibility. This method is accomplished by recursively adding a modified input adaptive schedule into an existing filtering system. When applied in sudden illumination changes, the performance is noticeably better than previous approaches.

Martins *et al.* [36] developed a novel classification mechanism that blends colour space discriminating skills with hysteresis and a dynamic learning rate to update the background model with sudden illumination change. Each channel element L^*, a^*, b^* is studied on its own, and the decisions obtained from each are merged using the AND rule, which produces good results compared to those obtained from a conclusion based on majority voting. To prevent noisy pixels, the colour distance of which is very significant to the decision threshold, from inadvertently altering the classification, a hysteresis method has been implemented. The dynamic background learning rate denoted by α_{DBG}, depends on the number of Gaussians present in the mixture, marked by M, such that $\alpha_{DBG} = M \cdot \alpha_{BG}$, where is a fixed minimum value for the learning rate of the background if the pixel is categorized as background. This method ensures that the model adaptation is performed more quickly in dynamic regions and more slowly in static regions. In contrast, if the pixel categorization shifts from foreground to background, a higher learning rate α_{UBG} is applied. This method encourages rapid adaptation when the background is shown, which helps prevent the appearance of phantom pictures.

Agrawal and Natu [37] developed GMM with BLOB analysis of the interconnected parts, including labelling and morphological operations, to increase the accuracy of foreground detection. The suggested model can be broken down into two stages: the training phase, which is responsible for producing a referential image, and the testing phase, which is in charge of producing a binary mask. This model determines the difference between the frame of reference $BMG(x, y)$ and the current, and then it applies a threshold value in order to extract the region of interest. When using this approach to construct the foreground model, a threshold value determined using the standard deviation was selected for each pixel. Compared to other methods, the results demonstrate that the concept of integrating GMM with blob analysis and morphological operation obtained a lower number of incorrectly classified pixels.

Su [38] presented a GMM with data model optimization to adapt light transitions. Calculating the gradient picture of the stream is the first step in the approach. The scar operator is used for this calculation. Then, integrate RGB and gradients and use figurative ways to eliminate noisy movement areas and combine those which remain. In order to reduce the likelihood of making an incorrect diagnosis, a comparison is made between the two models' outputs to arrive at a final make-up area. Ultimately, they conducted the assessment and comparison using three separate image streams. The results reveal this method increases the accuracy of the detection process by minimizing the occurrence of erroneous detection areas caused by sudden illumination changes.

3.3. Shadow

Foreground items frequently have shaded areas due to changing light, which typically affects the segmentation of foreground items and the execution of subsequent modules of an algorithm that models the background. More specifically, there is a substantial difference in the lighting, but just a minor variation in the colour, in a darkened area. A pixel is considered a component of the shadow in the scene if it is a part of the background model that has been made darker by a shadow produced by another object in the scene. Therefore, a reliable method should include this technique to eliminate shadows cast by the foreground regions or disregard shadows that aren't relevant to the problem at issue [39].

Yadav and Xiaogang [34] presented the innovative hybrid method relying on the GMM, subtraction of background, hue-saturation-value (HSV) colour model, feature extraction, and neural networks. In order to provide a clean foreground, the shadows cast by moving objects are identified, and then eliminated using the HSV colour model, and morphological operations are carried out. After the algorithms indicate that the detection is finished, the background is modified to conform to the dynamic background. Following the identification of objects, the shape properties of those objects are retrieved using Hu's seven-moment invariants of the training samples of the image data. These recovered shape features are then fed into the back propagation neural network (BPNN) during the training process. The system could erase the shadow's influence and accurately detect motion. The findings of the experiments have shown that the suggested approach has strong resilience as well as real-time performance in realistic environments.

Jin *et al.* [40] presented an Improved GMM-based automatic segmentation method (IGASM) by improving the approach for updating the background so that it can more effectively segment the floats on the sea surface. Following the mapping of the GMM's findings into an HSV colour space, a light-shadow classifier function is implemented to address the problems associated with shadows. After that, a morphological method is applied in order to refine the foregrounds that were previously acquired. In the end, the graph cuts technique is utilized to optimize the segmentation outcomes based on the spatial information in the video images. It is possible to detect stationary floats by improving the updating approach successfully. In order to solve the issues that are brought about by shadows in the segmentation results, a shadow discriminant function is utilized. Smoothing out the contour is accomplished through the open operation and the graph cuts algorithm. As a result, the accuracy of the segmentation results can be increased even more. The results of the experiments reveal that this method demonstrates a higher efficiency in the surface floats segmentation on the water, particularly in situations where there is a considerable shift in light. Still, the surface floats on the water do not move.

Lin and Chen [41] provide a method for the detection of moving objects that are based on GMM and visual saliency maps. This method can eliminate the disruption caused by the shadow situation and accomplish stable detection of moving objects. The researcher uses the GMM approach to construct models for the video sequences and then obtains the rough foreground objects. The foreground, however, has a significant amount of incorrect detection, and as a result, they cannot adequately extract the moving objects. In the second step of the process, which is refining the crude foreground objects, make use of the vision saliency to achieve reliable detection results. Following the conversion of each image frame to the L^*, a^*, b^* colour space, the L^*, a^*, b^* channels are smoothed using a Gaussian filter in order to remove small texture features as well as noise. Then, estimate the saliency maps for each channel and linearly merge those maps to generate the final saliency map. In the end, the final saliency maps and the foregrounds are combined in order to produce the moving objects. The shadow issue can be successfully addressed and resolved using this method.

3.4. Multimodal background

A multimodal situation is created whenever there is motion occurring in the background. The scenery in the background might have some motion, such as a fountain, the movement of clouds, the swaying of nearby trees, and a wave in the water. This movement can be regular or irregular at various intervals. The traditional GMM is capable of handling multimodal backgrounds robustly. However, the parameter K is fixed experimentally, and its value remains unchanged. This method is not ideal regarding the time required for detection, and computation [24]. An effective GMM improvement should be able to detect regular, or irregular motion [39].

Ou *et al.* [42] proposed an adaptive GMM (AGMM) with BPNN to extract the foreground objects in multimodal background conditions. In most cases, AGMM is employed to accomplish the twin goals of simplifying the algorithm and enhancing its precision. All image pixels can fit with the hybrid model if more single-Gaussian models exist. The neural network can figure out the statistical parameters of the image's noise, the model can alter the noise very well, and the foreground objects can be preserved entirely. This method solves the problem of defective foreground objects due to morphological processing. It eliminates the need for the model structure to make the trade between foreground objects and the noise. Because of this, they are implementing an adaptive version of the GMM supported by a neural network that can improve the robustness and performance of the entire system.

Zuo *et al.* [33] proposed a GMM-based technique for moving picture target recognition that overcomes multimodal backgrounds. Compared to other methods, this method's main advantage is that it eliminates disturbance from dynamic backgrounds and improves algorithm detection. The image sequence is initially used to divide the image of the video sequence. This method is followed by replacing each pixel value with the overall average of each pixel of the image block in the background modelling step. Then, the image block of the GMM mean approach is applied. The combination of mathematical morphology and a semisoft threshold function is employed in this article to remove noise from an image of the foreground detection during the process of detecting moving targets. The soft and hard threshold methods have distinct disadvantages. The researcher presents a revised semisoft threshold function to overcome the limitations of both the hard and soft threshold functions. The conventional GMM approach cannot update the background in real time, resulting in a ghosting and poor moving target identification accuracy. The adaptive background update algorithm uses the existing frame of image detection and the background model to tackle the problem. This quantitative and qualitative approach is preferable to the method that substantiates the claims made on the system's efficiency and adaptability.

3.5. Camouflage

The camouflage means the foreground may contain an element with a colour identical to the background. It creates confusion during the process of detection and makes it difficult to determine whether something is in front of the background. Even while the traditional GMM has fine-tuned crucial parameters, it still tends to create more false alarms.

Zhang *et al.* [43] use a combination of foreground matching with a measure of short-term stability for camouflage. The priority is given to matching probable foreground in incoming pixels by foreground models that have been constructed and updated using the foreground pixels that have been detected. During this time, the stability at the pixel level is assessed to ensure that an integrated foreground will be identified when a dynamic foreground procedure is being carried out. Suppose the currently-used pixel value does not match any pre-existing background models. In that case, the foreground model will be constructed with the currently-used pixel value acting as the model's mean and an enormous value working as the model's variance. Foreground models are given precedence over background models by this approach. Foreground models always try to match the pixels that come after them, which helps to lessen the likelihood that foreground pixels will not accurately represent the background models. Compute the short-term stability by utilizing the previous pixel only if the existing foreground model does not match the next pixel can be written as (10)

$$S = \frac{M \sum_{j=1}^{M} x_{t+j}^2 - \left(\sum_{j=1}^{M} x_{t+j}\right)^2}{M(M-1)} \tag{10}$$

where M are the total frame numbers representing the amount of time that has passed. The GMM's tolerance to camouflage at a slow pace is significantly improved by the combination of foreground matching and short-term stability measures.

Lima *et al.* [44] proposed a method for estimating a threshold for each region by feedback step. Spatial analysis is used to determine a threshold, which is then used to classify the pixel. In order to correct some classification mistakes, the segmentation is filtered before the threshold estimation is performed. This filtering technique removes disturbances and combines the entire area into one cohesive whole. During the feedback stage, the fixes are used to estimate the thresholds for the subsequent iteration. Since the filtering stage only corrects foreground mistakes, the improvement is concentrated on the vehicle areas. This strategy that has been recommended promotes the segmentation of regions that have already been segmented. As a result, the estimation of the threshold is comparable to a first-order Markov chain. This technique lowers the amount of error on subsequent iterations when applied to fixed areas.

3.6. Slow moving object

Motions with unique patterns, such as moving slowly or staying, are significant for security protection since they make detection exceedingly difficult [45]–[49]. In addition, these kinds of motions make it more challenging to identify perceived risks. In the actual world, many things move at a slower rate. When dealing with the kind of objects, the majority of the currently available GMM-based algorithms typically extract fragmented bodies. This results in a significant reduction in detection precision. The traditional GMM algorithm is predicated on the idea that the background is more likely to be seen than any particular foreground. Because of this assumption, there is a chance that a long-observed object that moves slowly would be misidentified as background. This might lead to significant complications. It is not appropriate to include a slow-moving object in the background until after the object has stopped moving altogether and has been still for a sufficient time [43].

Zhang *et al.* [43] use the foreground matching technique to detect slow-moving objects. GMM is used for the foreground and is continuously updated using newly identified foreground pixels. It takes into explicit consideration the spatial continuity of the objects that are moving. The foreground model is prioritized to fit the pixels, and pixels that provide the foreground models are explicitly designated as foreground. This is done to prevent the misleading inclusion of long-term foreground into the background. The results show progress has been made in improving the performance of the identification of slow-moving objects in complicated circumstances.

Fu *et al.* [50] studied an initial Gaussian background model (IGBM) based on an extended Kalman filter to improve GMM performance for the slow-moving object. In order to find a solution to this issue, the IGBM is not only incorporated with the GMM but also adds an extended Kalman filter [51]–[53] based tracker into the system whenever the grey value of this static object is comparable to the grey background value. Doing things this way can keep the moving item in its static condition indefinitely. Meanwhile, the tracker implemented in the system can reliably determine whether the grey of a stationary object is comparable to the grey background value. This research assumes a total of K single Gaussian distributions in the GMM. Hence, they can construct a new background model using these initial Gaussian distributions, which are IGBM. Due to the fact that it is made up of complete single Gaussian distributions rather than the initial GMM, it can keep the most abundant information from the model that was used to create the original background. The results indicate that the EGMM approach can efficiently address the challenge of recognizing moving objects that stop and start intermittently.

Zhang *et al.* [54] presented the GMM with confidence measurement (GMMCM) as a potential solution to the issue that the background subtraction model is susceptible to being easily contaminated by vehicles either moving slowly or temporarily stopped. The confidence measurement (CM) is then applied to each background model pixel to quantify each background pixel's current trust values. In order to avoid contamination of the background model in complicated urban traffic scenes, a technique for maintaining a balance between the dynamic changes in the brightness and the background, which makes use of a self-adaptive learning rate to keep the background model, has been developed. The result reveals that the proposed GMMCM does a great job of coping with cars that are going slowly or temporarily stopped.

3.7. Bootstrapping

Background initialization for bootstrapping video sequences is a technique that is frequently employed in intelligent video surveillance systems for the purpose of monitoring crowded public spaces. Background removal approaches have recently been concentrating on background initialization for bootstrapping video sequences. This is because a training duration devoid of foreground items is not always accessible in congested contexts [55], [56]. To put it another way, the definition of background initialization for bootstrapping video sequences is to estimate a background frame that does not contain any foreground items, given a video sequence that was captured by a stationary camera and in which the background is occluded by some foreground objects in each frame of the video sequence [57], [58]. This circumstance occurs when the objects in the foreground constantly occupy the background. In a situation like this, there are two conceivable outcomes. In the first scenario, the background is excessively active, and no single frame will be available that is devoid of any objects. The second scenario has a crowded background, but there are still some beginning frames that do not include any objects.

Amintoosi *et al.* [59] proposed the identification of the background through the use of the GMM with the QR decomposition method in linear algebra. R-values derived via QR decomposition can be used to deconstruct a given system to reflect the level of relevance possessed by each system's components that have been decomposed. After dividing the image into several smaller blocks, the researchers look at each to determine which makes a minor significant contribution to the overall picture and then choose those blocks. The simulation findings indicate that the background detection performance is superior to other methods.

Harville *et al.* [60] used a method for modelling the background that uses per-pixel, time-adaptive GMM in the combined input space of depth and luminance-invariant colour. In this method, the GMM is used to model the background. This combination is already novel, but the researchers make it even more so by introducing the concepts of modulating the background model learning rate based on scene activity and making colour-based segmentation criteria dependent on depth observations. Both of these ideas serve to improve the original combination further. These results demonstrate that the method has significantly larger robustness to troublesome phenomena than the previous state-of-the-art without sacrificing real-time performance. This makes it well-suited for a wide range of practical applications in video event detection and recognition. Table 1 summarizes all the background challenges in motion detection using GMM approaches.

Table 1. Summary of background subtraction challenges in motion detection using GMM

No.	Background issues	Modification of GMM	Authors
1	Image or video noise	1.MOD-AT based on adaptive threshold.	Luo *et al.* [27]
		2.Calculating the L2 Norm.	Wei and Zheng [28]
		3.Image block averaging method, wavelet semi-threshold, and adaptive background updating method.	Zuo *et al.* [33]
		4.Fuzzy distance in EM and FCM algorithm.	Kalti and Mahjoub [32]
		5.A multi-dimensional Gaussian Kernal density transform.	Chen and Ellis [29]
2	Sudden illumination change	1.Online dynamic learning.	Chen and Ellis [35]
		2.Colour space discrimination capabilities with hysteresis and dynamic learning rate.	Martins *et al.* [36]
		3.Blob analysis and morphology.	Agrawal and Natu [37]
		4.Data model optimization.	Su [38]
3	Shadow	1.HSV colour model features extraction and neural network.	Yadav and Xiaogang [34]
		2.Background updating strategy.	Jin *et al.* [40]
		3.Visual saliency map.	Lin and Chen [41]
4	Multimodal background	1.AGMM and BPNN hybrid method.	Ou *et al.* [42]
		2.Image block average method, wavelet semi-threshold, adaptive background updating method.	Zuo *et al.* [33]
5	Camouflage	1.Foreground matching and short-term stability measure.	Zhang *et al.* [43]
		2.The threshold for each region by feedback step.	Lima *et al.* [44]
6	Slow moving object	1.Foreground matching.	Zhang *et al.* [43]
		2.Initial GMM and Extend Kalman Filter.	Fu *et al.* [50]
		3.Confidence measurement.	Zhang *et al.* [54]
7	Bootstrapping	1.QR decomposition.	Amintoosi *et al.* [59]
		2.Learning rate adaptation based on scene activity.	Harville *et al.* [60]

4. CONCLUSION

The goal of this study is to have a comprehensive assessment of the numerous GMM methods that have been developed to address various background issues. In addition to that, it offers a concise description of GMM methods. Image or video noise, sudden illumination change, shadow, multimodal background, camouflage, bootstrapping and slow-moving object are some of the background difficulties that various GMM versions can address. Image and video noise are the difficulties that receive significant attention from researchers, while camouflage receives less of their focus. Different variants of GMM are capable of simultaneously addressing various challenges. This study assists researchers in selecting the correct version of GMM for their applications based on the study's findings. Moreover, GMM techniques, including comprehensive bibliography material, can provide helpful insights into this critical background topic and motivate for future research.

ACKNOWLEDGEMENTS

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] K. Goyal and J. Singhai, "Review of background subtraction methods using Gaussian mixture model for video surveillance systems," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 241-259, 2018, doi: 10.1007/s10462-017-9542-x.

[2] T. Bouwmans, F. El Baf and B. Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent patents on computer science*, vol. 1, no. 3, pp. 219-237, 2008, doi: 10.2174/1874479610801030219.

[3] T. Liu, X. Cao, and J. Jiang, "Visual object tracking with partition loss schemes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 633–642, 2016, doi: 10.1109/TITS.2016.2585663.

[4] T. S. Waykole and Y. K. Jain, "Detecting and tracking of moving objects from video," *International Journal of Computer Applications*, vol. 81, no. 18, pp. 0975–8887, 2013, doi: 10.5120/14224-2410.

[5] Y. Dai, Y. Wu, F. Zhou and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9813–9824, 2021, doi: 10.1109/TGRS.2020.3044958.

[6] M. Devanne, S. Berretti, P. Pala, H. Wannous, M. Daoudi and A. Del Bimbo, "Motion segment decomposition of RGB-D sequences for human behaviour understanding," *Pattern Recognition*, vol. 61, pp. 222–233, 2017, doi: 10.1016/j.patcog.2016.07.041.

[7] X. Y. Zhang, H. Shi, C. Li, K. Zheng, X. Zhu and L. Duan, "Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision," In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, pp. 9227–9234, 2019, doi: 10.1609/aaai.v33i01.33019227.

[8] D. S. Pham, O. Arandjelović and S. Venkatesh, "Detection of dynamic background due to swaying movements from motion feature," *IEEE Transactions on Image Processing*, vol. 24, no.1, pp. 332-344, 2014, doi: 10.1109/TIP.2014.2378034.

[9] Y. Xin, J. Hou, L. Dong and L. Ding, "A self-adaptive optical flow method for the moving object detection in the video sequences," *Optik*, vol. 125, pp. 5690–5694, 2014, doi: 10.1016/j.ijleo.2014.06.092.

[10] P. Han, J. Du, J. Zhou and S. Zhu, "An object detection method using wavelet optical flow and hybrid linear-nonlinear classifier," *Mathematical Problems in Engineering*, vol. 2013, pp. 1–14, 2013, doi: 10.1155/2013/965419.

[11] R. M. Baby and R. R. Ahamed, "Optical flow motion detection on Raspberry Pi," In *Fourth International Conference on Advances in Computing and Communications*, 2014, pp. 151–152, doi: 10.1109/ICACC.2014.42.

[12] D. Archana and S. Sanjeevani, "Moving object detection using optical flow and HSV," In *Evolution in Signal Processing and Telecommunication Networks*, vol. 839, 2022, pp. 49-55, doi: 10.1007/978-981-16-8554-5_6.

[13] Y. H. Cheng and J. Wang, "A motion image detection method based on the inter-frame difference method," *Applied Mechanics and Materials*, vol. 490–491, pp. 1283–1286, 2014, doi: 110.4028/www.scientific.net/AMM.490-491.1283.

[14] M. Wan, G. Gu, E. Cao, X. Hu, W. Qian, and K. Ren, "In-frame and inter-frame information-based infrared moving small target detection under complex cloud backgrounds," *Infrared Physics & Technology*, vol. 76, pp. 455–467, 2016, doi: 10.1016/j.infrared.2016.04.003.

[15] J. Guo, J. Wang, R. Bai, Y. Zhang and Y. Li , "A new moving object detection method based on frame-difference and background subtraction," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 242, o. 1, p. 012115, doi: 10.1088/1757-899X/242/1/012115.

[16] Y. Chen and J. Dong, "Target detection based on the interframe difference of block and graph-based," In *Proceedings of the 2016 9th International Symposium on Computational Intelligence and Design*, 2016, vol. 2, pp. 467–470, doi: 10.1109/ISCID.2016.2115.

[17] A. Akula, N. Khanna, R. Ghosh, S. Kumar, A. Das, and H. K. Sardana, "Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences," *Infrared Physics & Technology*, vol. 63, pp. 103–109, 2014, doi: 10.1016/j.infrared.2013.12.012.

[18] M. Kaushaland and B. S. Khehra, "BBBCO and fuzzy entropy-based modified background subtraction algorithm for object detection in videos," *Applied Intelligence*, vol. 47, no. 4, pp. 1008–1021, 2017, doi: 10.1007/s10489-017-0912-5.

[19] R. Kalsotra and S. Arora, "Background subtraction for moving object detection: explorations of recent developments and challenges," *The Visual Computer*,vol.38, pp. 4151-5178, 2022, doi: 10.1007/s00371-021-02286-0.

[20] A. F. Abbas, U. U. Sheikh, F. T. AL-Dhief and M. N. H. Mohd, "A comprehensive review of vehicle detection using computer vision," *TELKOMNIKA(Telecommunication Computing Electronics and Control)*, vol. 19, no. 3, pp. 838–850, 2021, doi: 10.12928/TELKOMNIKA.v19i3.12880.

[21] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection-a systematic survey," *Recent Patents on Computer Science*, vol. 4, no. 3, pp. 147–176, 2011, doi: 10.2174/1874479611104030147.

[22] X. Yu, X. Chen and H. Zhang, "Accurate motion detection in dynamic scenes based on ego-motion estimation and optical flow segmentation combined method," In*2011 Symposium on Photonics and Optoelectronics (SOPO)*, 2011, pp. 1-4, doi: 10.1109/SOPO.2011.5780637.

[23] M. Shah, J. D. Deng and B. J. Woodford, "Video background modeling: recent approaches, issues and our proposed techniques," *Machine Vision and Applications*, vol. 25, no. 5, pp. 1105–1119, 2014, doi: 10.1007/s00138-013-0552-7.

[24] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 1999, vol. 2, pp. 246–252, doi: 10.1109/CVPR.1999.784637.

[25] D. S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 5, pp. 827–832, 2005, doi: 10.1109/TPAMI.2005.102.

[26] Y. Zhao and Y. Su, "Vehicles detection in complex urban scenes using Gaussian mixture model with FMCW radar," *IEEE Sensors Journal*, vol. 17, no. 18, pp. 5948–5953, 2017, doi: 10.1109/JSEN.2017.2733223.

[27] 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, doi: 10.3390/ijgi10110742.

[28] H. Wei and W. Zheng, "Image denoising based on improved Gaussian mixture model," *Scientific Programming*, vol. 2021, pp. 1–8, 2021, doi: 10.1155/2021/7982645.

[29] Z. Chen and T. Ellis, "A self-adaptive Gaussian mixture model," *Computer Vision and Image Understanding*, vol. 122, pp. 35–46, 2014, doi: 10.1016/j.cviu.2014.01.004.

[30] Z. Chen, N. Pears, M. Freeman and J. Austin, "Background subtraction in video using recursive mixture models, spatio-temporal filtering and shadow removal," in *International Symposium on Visual Computing*, vol. 5876LNCS, no. PART 2, 2009, pp. 1141–1150, doi: 10.1007/978-3-642-10520-3-109.

[31] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002, doi: 10.1109/34.1000236.

[32] K. Kalti and M. A. Mahjoub, "Image segmentation by gaussian mixture models and modified FCM algorithm," *International Arab Journal of Information Technology*, vol. 11, no. 1, pp. 11–18, 2014.

[33] J. Zuo, Z. Jia, J. Yang and N. Kasabov, "Moving target detection based on improved Gaussian mixture background subtraction in video images," *IEEE Access*, vol. 7, pp. 152612-152623, 2019, doi: 10.1109/ACCESS.2019.2946230.

[34] B. K. Yadav and J. Xiaogang, "An Algorithm for intelligent identification of moving objects in natural environment," *International Journal of Machine Learning and Computing*, vol. 10, no. 5, pp. 637–647, 2020, doi: 10.18178/ijmlc.2020.10.5.985.

[35] 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)*, 2011, pp. 1769-1776, doi: 10.1109/ICCVW.2011.6130463.

[36] 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, 2018, doi: 10.1007/s10044-018-0699-y.

[37] S. Agrawal and P. Natu, "An improved Gaussian mixture method-based background subtraction model for moving object detection in outdoor scene," In *2021 4th International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2021, pp. 1–8, doi: 10.1109/ICECCT52121.2021.9616883.

[38] Y. Su, "Target detection algorithm and data model optimization based on improved Gaussian mixture model," *Microprocessors and Microsystems*, vol. 81, pp. 103797, 2021, doi: 10.1016/j.micpro.2020.103797.

[39] Y. Xu, J. Dong, B. Zhang and D. Xu, "Background modelling methods in video analysis: a review and comparative evaluation," *CAAI Transactions on Intelligence Technology*, vol. 1, no. 1, pp. 43–60, 2016, doi: 10.1016/j.trit.2016.03.005.

[40] X. Jin, P. Niu, and L. Liu, "A GMM-based segmentation method for the detection of water surface floats," *IEEE Access*, vol. 7, pp. 119018-119025, 2019, doi: 10.1109/ACCESS.2019.2937129.

[41] L. L. Lin and N. R. Chen, "Moving objects detection based on gaussian mixture model and saliency map," *In Applied Mechanics and Materials*, vol. 63-64, pp. 350–354, 2011, doi: 10.4028/www.scientific.net/AMM.63-64.350.

[42] X. Ou *et al.*, "Adaptive GMM and BP neural network hybrid method for moving objects detection in complex scenes," . *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 2, pp. 1950004, 2019, doi: 10.1142/S0218001419500046.

[43] C. Zhang, X. Wu and X. Gao, "An improved Gaussian mixture modelling algorithm combining foreground matching and short-term stability measure for motion detection," *Multimedia Tools and Applications*, vol. 79, no. 11-12, pp. 7049–7071, 2020, doi: 10.1007/s11042-019-08210-y.

[44] 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," In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, 2015, pp. 204–209, doi: 10.1109/BRACIS.2015.33.

[45] X. Cao, L. Yang and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," *IEEE Transactions on Cybernetics*, vol. 46, no. 4, pp. 1014–1027, 2015, doi: 10.1109/TCYB.2015.2419737.

[46] S. Puhan, D. K. Rout and N. K. Kamila, "Slow and fast moving object detection under illumination variation condition," *American Journal of Signal Processing*, vol. 3, no. 5, pp. 121–131, 2013, doi: 10.5923/j.ajsp.20130305.01.

[47] P. K. Sahoo, P. Kanungo and S. Mishra, "A fast valley-based segmentation for detection of slowly moving objects," *Signal, Image and Video Processing*, vol. 12, no. 7, pp. 1265-1272, 2018, doi: 10.1007/s11760-018-1278-9.

[48] B. N. Subudhi and P. K. Nanda, "Detection of slow-moving video objects using compound Markov random Field model," In *TENCON 2008-2008 IEEE Region 10 Conference*, 2008, pp. 1–6, doi: 10.1109/TENCON.2008.4766385.

[49] A. Utasi and L. Czúni, "Reducing the foreground aperture problem in mixture of Gaussians based motion detection," In *2007 14th International Workshop on Systems, Signals, and Image Processing and the 6th EURASIP Conference focused on Speech and Image Processing*, 2007, pp. 157–160, doi: 10.1109/IWSSIP.2007.4381177.

[50] H. Fu, H. Ma and A. Ming, "EGMM: an enhanced Gaussian mixture model for detecting moving objects with intermittent stops," In *2011 IEEE International Conference on Multimedia and Expo*, 2011, pp. 1–6, doi: 10.1109/ICME.2011.6012011.

[51] X. Kai, C. Wei, and L. Liu, "Robust extended Kalman filtering for nonlinear systems with stochastic uncertainties," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 2, pp. 399–405, 2009, doi: 10.1109/TSMCA.2009.2034836.

[52] C. Urrea and R. Agramonte, "Kalman filter: historical overview and review of its use in robotics 60 years after its creation," *Journal of Sensors*, vol. 2021, pp. 1-21, 2021, doi: 10.1155/2021/9674015.

[53] L. Torres, J. Jiménez-Cabas, O. González, L. Molina, and F. R. López-Estrada, "Kalman filters for leak diagnosis in pipelines: brief history and future research," *Journal of Marine Science and Engineering*, vol. 8, no. 3, pp. 173, 2020, doi: 10.3390/jmse8030173.

[54] 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)* , 2015, pp. 1-7, doi: 10.1109/IEMCON.2015.7344486.

[55] K. Toyama, J. Krumm, B. Brumitt and B. Meyers, "Wallflower: principles and practice of background maintenance," In *Proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 1, pp. 255-261, doi: 10.1109/ICCV.1999.791228.

[56] D. Farin, P. H. de With and W. Effelsberg, "Robust background estimation for complex video sequences," In *Proceedings 2003 International Conference on Image Processing*, 2003, vol. 1, pp. 1-145, doi: 10.1109/ICIP.2003.1246919.

[57] V. Reddy, C. Sanderson, B. C. Lovell and A. Bigdeli, "An efficient background estimation algorithm for embedded smart cameras," In *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2009, pp. 1-7, doi: 10.1109/ICDSC.2009.5289348.

[58] A. Colombari and A. Fusiello, "Patch-based background initialization in heavily cluttered video," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 926-933, 2009, doi: 10.1109/TIP.2009.2038652.

[59] M. Amintoosi, F. Farbiz, M. Fathy, M. Analoui and N. Mozayani, "QR decomposition-based algorithm for background subtraction," In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, 2007, vol. 1, pp. I-1093-I–1096, doi: 10.1109/ICASSP.2007.366102.

[60] M. Harville, G. Gordon and J. Woodfill, "Foreground segmentation using adaptive mixture models in color and depth," In *Proceedings IEEE workshop on detection and recognition of events in video*, 2001, pp. 3-11, doi: 10.1109/EVENT.2001.938860.

BIOGRAPHIES OF AUTHORS

Nor Afiqah Mohd Aris received her Bachelor of Science (Mathematics Technology) in 2019 and Master of Science (Mathematics) in 2021 from Universiti Tun Hussein Onn Malaysia, Malaysia, before pursuing a Doctor of Philosophy of Science (Mathematics) in 2022. She is currently a Graduate Research Assistant at the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Malaysia. She runs research in image processing for real-time traffic flow analysis supported by the Ministry of Higher Education (MOHE) Malaysia through the Fundamental Research Grant Scheme (FRGS). Her research interests include smart city, pattern recognition and machine learning. She can be contacted at email: norafiqahmohdaris@gmail.com.

Ts. Dr. Siti Suhana Jamaian received her first degree in Industrial Mathematics from Universiti Teknologi Malaysia in 2007. She then furthered her study with a Master's degree in the same university and graduated in 2009. In 2013, she completed her PhD in Applied and Computational Mathematics at the School of Mathematics, The University of Edinburgh. She is now a Senior Lecturer in the Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia. Her current research interests in applied mathematics include the chemostat model, nutrient removal, industrial wastewater, parameter estimation and traffic flow. She can be contacted at email: suhana@uthm.edu.my.