International Journal of Information Processing and Communication (IJIPC) Vol. 8 No. 1 [May, 2020], pp. 35-54
Online: ISSN 2645-2960; Print ISSN: 2141-3959

A Systematic Review of Background Subtraction Algorithms for Smart Surveillance System

Olayemi Mikail Olaniyi[1], Jibril Abdullahi Bala[2], Shefiu Olusegun Ganiyu[3], and Praise Emmanuel Wisdom[4]

[1,2,3,4]Federal University of Technology Minna, Niger State, Nigeria
E-mails: [1]mikail.olaniyi@futminna.edu.ng, [2]jibril.bala@futminna.edu.ng, [3]shefiu.ganiyu@futminna.edu.ng, [4]wisdom.praise@st.futminna.edu.ng

Abstract - Nowadays, the demand for video surveillance systems is rapidly increasing as an essential component of many security formations and apparatuses. Many times, video surveillance has demonstrated its importance and benefits by providing immediate supervision of environment, people and their valuable properties. However, surveillance has been a major issue in almost every facet of life since infrastructure, companies, homes and offices required the safety of their possessions from intruders. Therefore, this paper presents a comprehensive and systematic literature review on existing methods in motion detection such as background subtraction, as well as, an overview on basic concepts in surveillance technologies and the drawbacks of various methods. Consequently, in this paper an enhanced methodology to improve the existing method in smart surveillance systems is presented.

Keywords *– Surveillance, Motion detection, Image Processing, Background Subtraction, Gaussian Mixture Model.*

1. Introduction

Video surveillance systems plays a very important role in various disciplines of human endeavours, such as personal security, banking and business premises. Beginning from residential houses to massive industries, video surveillance is now crucial and vital to satisfy the safety aspects in numerous ways (Marx, 2015). Burglary and theft have always been a problem for normal residents, particularly for those living in the large cities of developing nations like Nigeria. Thus, it is rather indispensable to find an efficient way to significantly reduce such crimes. Generally, the world has been moving towards a "surveillance society" for the past fifty years (Lyon, 2001). Surveillance simply can be defined as the process of closely monitoring changing information like activities, conducts, behaviors for the justification or purpose of protecting, managing, influencing and directing people (Lyon, 2007). The area of surveillance is progressively becoming a topic of academic study (Marx, 2015), achieved through studies of research centers (Fujika, Plakninly and James,2017), books (Lyon, 2007) and peer-reviewed academic journals. According to Clapper in Ashish and Choubaey (2013): "*In the future, intelligence services might use the internet of things for identification, surveillance, monitoring, location tracking, and targeting for recruitment, or to gain access to networks or user credentials,*". The reality of this assertion is what manifests in modern day surveillance systems.

Insecurity has been a major challenge in recent times, from the loss of properties, to theft and robbery, to the loss of lives in most developing countries like Nigerian homes, particularly residents of big cities within the nation. The cost of the standard Close Circuit Television (CCTV) is too high and still does not present the

Online: ISSN 2645-2960; Print ISSN: 2141-3959

facilities to improve surveillance in cases where the user is not within the vicinity of the installed surveillance system. The ability to monitor and survey homes even in far distance is what actually led to the term *"surveillance"* (Lyon, 2007).

Primarily, this study is premised along a systematic review of existing literatures to evolve an enhanced methodology for smart surveillance system. Precisely, the method proposed in the study uses an adaptive Gaussian mixture technique to improve the basic background subtraction technique in motion detection. This paper is intended to appraise various techniques employed in the development of a smart surveillance systems using the camera for motion detection. Improper segmentation of foreground especially in overlapped images is prominent in the basic background subtraction technique. The use of an adaptive Gaussian Mixture Technique is proposed to enhance the contemporary Background Subtraction technique using computer vision (camera) for both surveillance and motion detection.

The remaining sections of this paper are organized as follows: Section 2 presents background of some fundamental concepts of smart surveillance, Section 3 reviews related literatures, Section 4 presents the proposed methodology and Section 5 concludes the paper.

2. Background of Related Fundamental Concepts

This section reviews background concepts of different Background Subtraction Techniques used for object detection. Some of the techniques used are: frame differencing, approximate median, running average, and running Gaussian average. There are issues faced by many researchers in attempting to obtain an efficient background subtraction. Some of these issues are loss of smoothness, effects due to noise, maintaining robustness and improper segmentation of foreground object. Authors in Cheng, Gong, Schuurmans, and Terry (2011) outlined related works on background subtraction with the issues that are peculiar to each method. They obtained most of these issues by introducing a series of discriminative learning algorithms and subsequently compared their results with other works. However, the buffer size used by the researchers is limited in order to maintain constant speed. This section also presents and critically examines the adopted techniques in other to improve the existing background subtraction techniques.

2.1 Computer Vision

Computer vision complements methods and techniques through which artificial vision systems can be constructed and employed in a reasonable way for practical applications. This area of computing includes the software, hardware and imaging techniques necessary for various methods of image processing (Davis, 2005). Basically, the operation of a computer vision system is composed of the following stages: image acquisition and image processing.

2.1.1 Image Acquisition

This is the transfer of an electronic signal from the sensor to the numerical representation with a device such as a camera (Zareiforoush, Minaei, Alizadeh, and Banakar, 2015). Mainly, two types of image acquisition camera are distinguished by the shape of the scanning cameras used (the area or path). On one hand, conventional or area-scan camera produces images in each cycle of exposure. On the other hand, the line scanning camera captures only one row of pixels at a time. For a two-dimensional image acquisition, it is necessary to move the object to be captured by using a conveyor or moving the camera along a stationary object. The quality of images obtained by a computer vision system is directly affected by the lighting used during the acquisition phase. In this way, all the efforts invested in the use of appropriate lighting will improve the performance and reliability of the system and reduce the complexity of the software used in the processing stage (Hornberg, 2017).

2.1.2 Image Processing

This involves tasks undertaken to manipulate digital images for the purpose of improving the image quality, reducing noise or correcting exposure problems. In addition, image processing technique refers to the process of distinguishing one region from another, for the purpose of extracting information. Image processing can be divided into low-level, mid-level and high-level processing. Low-level processing, or preprocessing, includes operations for grayscale adjustment, correction of focus, contrast or sharpness enhancement and noise reduction. These operations generate a new image and work to improve the quality of the image or to change the position of the object of interest with geometric transformations (Hornberg, 2017).

Meanwhile, mid-level process involves the operation of segmentation (partition of the image in the area), description and classification of objects present in the image. Image segmentation produces a set of contours or region. Extraction of attributes that characterizes the area produced by segmentation is required to evaluate a series of characteristics of the regions of interest. For example, the ellipse parameters allow the determination of the orientation and size of the area. Bounding box allows the calculation of the height and width of the region of interest (Schaeffel & Hornberg, 2017). Filters such as Sobel, Laplacian and Laplacian of Gaussian can be used to determine which area of sudden changes in intensity occur in the image. Gabor filters selectively separate into elements in an image within a specific distance orientation and frequency. The purpose of edge detection is to generate a binary image in which the values of zero indicate an edge in the image. The detector may optionally also return other information such as scale and orientation relative to the edge. Some examples are Canny edge detector, Harris and The Scale Invariant Feature Transform (SIFT) SIFT. SIFT detector is a method to identify points of interest. It associates scale and orientation information for each point resulting from the detection process (Prince, 2012). High-level process, or image classification, involves the recognition and classification of the region of interest which is usually done by statistical or neural network classifier.

2.2 Background Subtraction Techniques

Background Subtraction technique involves subtracting the background from a video stream to obtain a foreground object and determine if motion is detected on the area of focus of the camera. Various techniques have been employed to improve the output of the background subtraction. Thus, eight of the well-known techniques are explained subsections 2.2.1 to 2.2.8:

2.2.1 Frame Differencing

Frame differencing (FD) is the most basic techniques or procedures in Background Subtraction (BGS). FD involves finding the absolute difference in between the frame and the previous image background or frame (Ashish, Choubaey, Michin, & Funkar, 2019). The absolute difference is then compared with the corresponding threshold value, A, to detect objects as shown in equation (1),

$$Fg_i(x,y) = \begin{cases} 1 & |F_i(x,y) - B_i(x,y)| > A \\ 0 & Otherwise \end{cases} \quad (1)$$

Where F_i is the current frame intensity value, B_i is the background intensity value. This technique uses the background frame for all video sequences.

2.2.2 Approximate Mean Filter

Approximate median (AM) is an adaptive, dynamic, non-probabilistic, and intuitive technique obtained by calculating the difference between two frames of video and using this difference in determining the perfect method for updating the background (Singh *et al.*, 2009). Really, AM is regarded as one of the most acceptable methods in background subtraction, because it provides the most accurate pixel identification. Several studies such as Ahmed *et al.*, (2002); Mohammed *et al.*, (2002); Ashish & Choubaey (2013); Barnerjee, (2014) and Harum *et al.,* (2018) have evaluated the effectiveness of the AM filter algorithm. Similarly, He *et al.,* (2008) studied the effectiveness of the Approximate median algorithm as part of their improved and optimized algorithm for vehicle detection in an embedded system. Their approach produced accurate results with a lower computational time when detecting and tracking vehicles in a traffic scene. Equation (2) presents an updated AM in the reference or background frame of each video sequence. The successful background frame, depends on the intensity values of both the current frame and the background frame.

$$B_{i+1}(x,y) = \begin{cases} F_i(x,y)+1 & |F_i(x,y) > B_i(x,y)| \\ F_i(x,y)-1 & Otherwise \end{cases} \quad (2)$$

2.2.3 Running Average

The Running Average (RA) is another approach for updating the background image. A particular pixel is classified or categorized as background when the pixel values belong to the appropriate distribution of the background model, and if otherwise, the mean of the distribution is updated (Park, Tabb, & Kak, 2006). Images that are updated later are used in the scene changes. The computational effect in the background running average is lower because they compute only the weighted sum of the two images, and thus, it produces a low computational time and minimized space (Zheng & Fan, 2010). Several researchers have used this technique to detect moving objects in videos captured by a static camera (Harum,Ali,Zakaria, and Anawar,2018);(Ashish and Choubaey, 2013); (Ashish , Choubaey, Michin, and Funkar, 2019); (Kato, Joga ,Rittscher, and Blake, 2002). Equation (3) shows a specified learning rate based on the previous background frame, where α signifies the learning rate and A is the threshold value:

$$B_{i+1}(x,y) = \begin{cases} B_i(x,y) & |F_i(x,y) > A| \\ \alpha * F_i(x,y) + (1-\alpha) * B_i(x,y) & Otherwise \end{cases} \quad (3)$$

2.2.4 Running Gaussian Average

This technique employed a combination of both Gaussian function and running average. Generally, the running Gaussian Average (RGA) has significant advantages over other approaches due to the fact that the processing time is lower and takes advantage of lower memory compared with a non-recursive such as mixture of Gaussian (MOG) and the density of the kernel estimation (KDE) (Chen, 2008). Equation (4) depicts how the frame of reference image is represented by the mean μ, which is sequentially updated using this method. Unlike the two previous techniques discussed AM and RA, this method uses a threshold value $2\sigma_i$

Online: ISSN 2645-2960; Print ISSN: 2141-3959

$$\mu_{i+1} = \alpha F_i + (1-\alpha)\mu_i$$
$$\sigma_{i+1}^2 = \alpha(F_i - \mu)^2 + (1-\alpha)\sigma^2 \quad (4)$$

2.2.5 Image Processing Technique in BST – Super pixel Extraction Technique

Super pixel is a polygonal part of any digital image or frame which is quite larger than a normal pixel but it is reduced in other similar properties such as brightness, texture, color, and vector (Ren and Malik, 2003).In general terms, super pixels are groups of pixels with approximately same pixel values. The overall aim of extracting the super pixels from an image or frame is to reduce the size of data that undergoes for the background subtraction, but other than this stated objective there are other relevancies to the usage of the super pixel extraction technique which are computational efficacy, less error rate, reduction in the number of primitives and theory (Mori, Ren, Efros, and Malik, 2004). Moreover, things to take note of during super pixel extraction are pixel collisions must never occur such that none of the pixels must belong to more than one super pixel. Now consider k as the total number of super pixels to be extracted from an image or frame the total number of pixels in the image is m, the value of the n^{th} pixel is represented by P_n, fix a minimum threshold $T_1 = 0$ and a maximum threshold $T_2 = \pm 1$ for each of the super pixel. The pixel value of the initial pixel in the superpixel is taken as the initial value for that super pixel and the difference between each pixel values are compared with that initial pixel value. Equation 5 shows how to determine the difference between each pixel in a super pixel.

$$P(K) = \sum_{n=0}^{m} P_n - P_1 \quad (5)$$

If the difference of pixel values lies between the threshold $T_1\ (zero)$ to $T_2(\pm one)$ then the pixel is considered to be in the same Super-pixel, else the pixel belongs to a different superpixel. The mathematical expression (6) below gives the overall behavior of superpixel extraction

$$P_n = \begin{cases} if\ T_1 \leq P_n \leq T_2\ then\ current\ superpixel \\ else\ neighboring\ superpixel \end{cases} \quad (6)$$

Every pixel in an image or frame undergoes the expression (6) and hence could achieve super pixel extraction. Also, the direction of the pixel that does not lie between the threshold is recorded with respect to the initial pixel in order to form a new super pixel in that direction. The superpixel algorithm is well expounded in (Yang, 2009) and a diagrammatical representation of the process is shown in Figure 1.

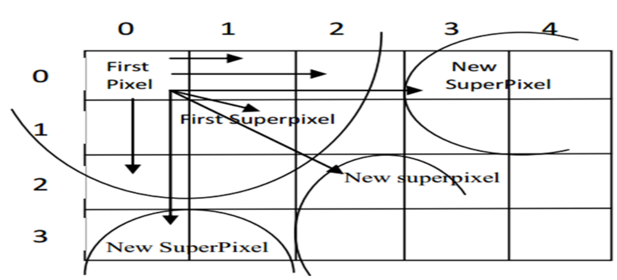

Figure 1: Superpixel Extraction of an image (Source: Yang, 2009)

Input: Image or frame of a video
Output: Image or frames with extraction of superpixel
Method:

1. Initialize the first pixel value P_1 as the initial value of the superpixel
2. Retrieve a pixel P_n
3. Compare P_n w ith P_1 i.e. by estimating $P_n - P_1$
4. If P_n falls between T_1 and T_2 then current superpixel, increment n then goto 2 else goto 5
5. Create next superpixel along the direction of P_n from the current superpixel
   a. Assign the threshold values
   b. Initialize the current pixel value as initial pixel value of this current superpixel
   c. goto 3
6. Stop if $P_n = P_m$

2.2.6 Image Processing Technique in BGS– Canny Edge Detection

The purpose of edge detection in general is to significantly reduce the amount of information in an image or picture, while maintaining the structural properties to be used for further image processing. Specifically, Canny, (1986) developed an algorithm that is now the most widely used standard and accepted algorithm for edge detection of an image/frame due to its optimal characteristics such as low error rate, response to single edge.(Mai, Hung, Zhong and Sze, 2008); (Azernikov, 2008). In Canny (1986), Canny edge detection is used for feature detection, feature extraction and to capture changes in an image or frame of a video in the background. The stated goal of edge detection is to remove or eliminate the irrelevant data to be processed. This algorithm operates on five (5) steps namely:

1. Smoothening
2. Finding gradients
3. Non – maximum suppression
4. Double thresholding, and
5. Edge tracking by hysteresis (Canny, 1986)

The Smoothening step is to reduce noise from the image or frame of the video because it is inevitable that any image retrieved from a camera not to contain some amount of noise. So, in order to prevent that noise from being mistaken to be an edge, the noise has to be reduced by applying the Gaussian filter.

The second or gradients finding step involves determining the gradients. Canny algorithm basically finds edges where the grayscale intensity of the image changes the most. These areas are found by determining gradients of the image. And in this case the gradient determination is applied to the superpixel not the pixel. The gradient of the image is determined using the Sobel- operator as suggested by Canny (Canny, 1986), which makes use of Euclidean distance to identify the magnitude. This method of applying gradient for superpixels reduces the computational complexity in a large fashion that is in terms of the number of pixels and parameters undergone for the gradient measure.

After gradient of superpixel is determined, the resulting image/frame shows the marked edges along the superpixel boundaries which may not be the actual object edge because a superpixel is a set of subpixels with similar properties. Thus, the obtained image will be of blurred edge (not sharp) due to large superpixel size. The sharp edges are obtained in the next step of Canny edge detection algorithm called non-maximum suppression.

Online: ISSN 2645-2960; Print ISSN: 2141-3959

The third step of non – maximum suppression involves identifying sharp edges from the blurred edges of the second step (Canny, 1986), whereby every subpixel of every superpixel in the blurred edges is undergone for non-maximum suppression such that edge strength of every subpixel is compared with the edge strength of pixels in the positive and negative gradients. Those pixels that have high local maxima and edge strength in this comparison are taken as edge pixels and others are deleted or suppressed. The output image will be marked with sharp edges but the weak edges are still not yet identified accurately due to noise or some rough surfaces.

The fourth step in the edge detection technique is the double thresholding and this is particularly useful to identify weak edges of the image or frame using three thresholding ranges. Canny uses double thresholding for grayscale images/frames but for color images we are going to employ the multiband or color thresholding, where the images must be first converted from RGB image to HSV and three threshold values are used (Bezdek, 1981). The most important issue to be considered in color thresholding is the selection of thresholding ranges due to the fact that the accuracy of output images/frames is also based on this threshold values. After identifying the weak edges, further analyses would reveal whether the selected weak edges belong to the edges of the object or the noisy pixels. This task is handled by the final step which is the edge tracking by hysteresis.

2.2.7 Image Processing Technique in BST – Fuzzy C means Technique

Fuzzy C means was first developed by Dun and improved by Bezdek (1981). It is the most efficient clustering algorithm for segmenting the image which allows a particular pixel to exist in more than one cluster and also to be able to identify the degree of membership in those clusters. This helps greatly in segmenting objects of an image efficiently and accurately (Banerjee, 2014). Fuzzy C means algorithm is used in Ahmed *et al.,* (2002) to segment the background and foreground object to perform background subtraction of consecutive frames of a video. This algorithm achieves segmentation of pixels whether belonging to the background frame or not. The main advantage of this method is, identifying the pixel membership in each cluster helps to segment the overlapped objects accurately.

One of the most important criteria to be considered in fuzzy c means is choosing the centroid. Initially the centroid is chosen based on the mean of the data points or randomly and the exact centroid can to calculated finally after several iterations of finding the membership degree of pixels. Fuzzy c-means is an important tool for grouping objects in image processing. To improve the accuracy of grouping noise in an image, the term spatial was introduced in 70s by mathematician (Ahmed *et al.* 2002,). Fuzzy c-means algorithm has also been used to distinguish different activities using image-based attributes (Barnerjee, 2014).

Application of k-means cluster algorithms in Image segmentation has been used for medical imaging, object detection and pattern recognition. Nevertheless, limitations such as shadows, noise and variations in camera, traditional hard groupings from image processing can be complicated in basic image processing tasks according to Yang (2009). From literature, Fuzzy c-means algorithm has proved to be more appropriate for performing these image processing tasks.

2.2.8 Gaussian Mixture Model

In machine learning, two main areas can be distinguished: supervised learning and unsupervised learning. The main difference between the two lies in the nature of the data as well as in the approaches used to process it.

Online: ISSN 2645-2960; Print ISSN: 2141-3959

Clustering is an unsupervised learning problem in which clusters of points in a dataset that share some common characteristics can be determined (Zivkovic, 2014). Having a dataset that looks like that in Figure 2.

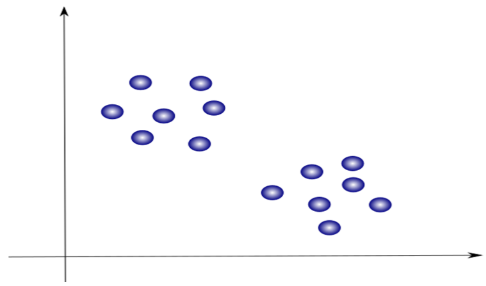

Figure 2: Graph showing grouping of points (Source: Zivkovic, 2014)

In order to find sets of points that seem close to each other, in this case, the system can clearly identify two groups of points that is colored respectively in blue and red Figure 3:

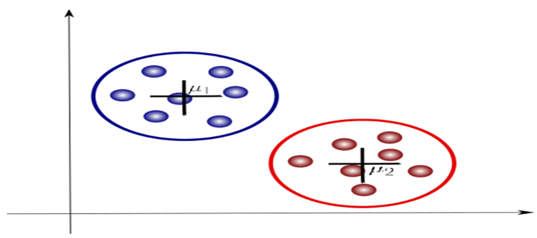

Figure 3: clusters being identified (Source: Zivkovic, 2014)

In Figure 3, μ1 and μ2 are the centroids of each cluster and are parameters to identify each of them. A popular classification algorithm known as K-means, used an iterative approach to update the parameters of each class. Specifically, it will calculate the averages (or centroids) of each cluster, then calculate their distance to each data point (Stauffer & Grimson, 1999). These are then labeled as part of the group identified by their nearest centroid. This process is repeated until a convergence criterion is satisfied, for example changes in cluster assignments cannot be seen (Zivkovic, 2014). An important feature of K-means is that it is a hard-clustering method, which means that it will associate each point with one and only one cluster. One of the limitations of this approach is that there is no measure or probability of uncertainty telling us how much a data point is associated with a specific cluster. This brings about the flexible clustering approach which is what Gaussian mixing model (or simply GMM) presented in (Zivkovic, 2014).

A Gaussian mixture is a function composed of several Gaussians, each identified by $k \in \{1, ..., K\}$, where K is the number of clusters in our dataset. Each k of Gauss in the mixture includes the following parameters: A mean μ that defines its center, a covariance Σ that defines its width. This would be equivalent to the size of an ellipsoid in a multivariate scenario, and a probability of mixing π that defines the size of the Gaussian function. Illustrating these parameters graphically we have the graph shown in Figure 4:

Online: ISSN 2645-2960; Print ISSN: 2141-3959

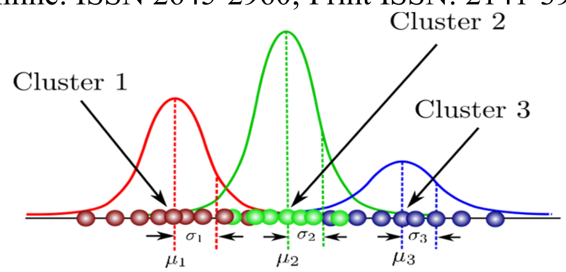

Figure 4: Clusters of Gaussian Mixtures (Source: Zivkovic, 2014)

Here it can be seen that there are three Gaussian functions, K = 3. Each Gaussian explains the data contained in each of the three available clusters (Harville, 2002). The mixing coefficients are themselves probabilities and must fulfill the condition in equation 7:

$$\sum_{k=1}^{K} \pi_k = 1 \qquad (7)$$

Now, to determine the optimal values for these parameters, each Gaussian corresponds to the data points belonging to each cluster. This is exactly what the maximum likelihood does (Kato, Joga, Rittscher, & Blake, 2002). In general, the Gaussian density function is given by the equation 8:

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right) \qquad (8)$$

Where x represents data point, D is the number of dimensions of each data point. μ and Σ are mean and covariance. Given a dataset consisting of N = 1000 three-dimensional points (D = 3), then x will be a matrix of 1000 × 3. μ will be a vector of 1 × 3, and Σ will be a matrix of 3 × 3. For further purposes, it is useful to take logs of this equation, which is given by equation 9:

$$\ln N(X|\mu,\Sigma) = -\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln \Sigma - \frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu) \qquad (9)$$

Introducing some additional notations to equation 9, to determine what is the probability that the data point x_n comes from Gaussian k, this is expressed as:

$$p(z_{n\,k} = 1|X_n) \qquad (10)$$

Which reads "given a data point x, what is the probability that it comes from Gaussian k?" In this case, z is a latent variable that only takes two possible values. This is one when x comes from Gaussian k, and zero otherwise. Actually, we cannot see this variable z in reality, but knowing the probability of occurrence will be useful in helping us determine the mixed Gaussian parameters, as discussed as follows (Ashish and Choubaey, 2013). Likewise, we can state the following:

$$\pi_k = p(z_k = 1) \qquad (11)$$

Online: ISSN 2645-2960; Print ISSN: 2141-3959

This means that the overall probability of observing a point that comes from k Gaussian is actually equivalent to the mixing coefficient for that Gaussian. This makes sense, because the larger the Gaussian, the more likely it is to be expected. Now let z be the set of all latent variables z possible, hence as shown in equation (12);

$$\boldsymbol{Z} = \{z_1, \dots, z_K\} \qquad (12)$$

Knowing in advance that each z occurs independently of the others and that they can only take the value of one when k equals the cluster from which the point originated.

$$p(\boldsymbol{z}) = p(z_1 = 1)^{z_1} p(z_2 = 1)^{z_2} \dots . p(z_K = 1)^{z_K} = \prod_{k=1}^{K} \pi_k^{z_k} \qquad (13)$$

Now, what about finding the probability of observing the data since they come from Gaussian k? It turns out that it's actually the Gaussian function itself. Following the same logic used to define p (z) we have:

$$p(X_n|\boldsymbol{z}) = \prod_{k=1}^{K} N(\boldsymbol{X_n}|\mu_k, \Sigma_k)^{z_k} \qquad (14)$$

All this is being done to determine what the probability z is from our observation x. Well, it turns out that the equations computed, as well as the Bayes rule, will help us determine that probability. From the product probability rule, knowing that

$$p(\boldsymbol{X_n}, \boldsymbol{z}) = p(\boldsymbol{X_n}|\boldsymbol{z})p(\boldsymbol{z}) \qquad (15)$$

The operands have been found, now to make use of the Bayes rule, we will first need p (xn) and not p (xn, z). getting rid of z we have:

$$p(\boldsymbol{X_n}) = \sum_{k=1}^{K} p(\boldsymbol{X}_n|\boldsymbol{z})p(\boldsymbol{z}) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{X}_n|\mu_k, \Sigma_k) \qquad (16)$$

Equation (16) defines a Gaussian Mixture, and clearly, it depends on all the parameters mentioned earlier to determining the maximum likelihood of the model. Therefore, finding the probability as the common probability of all observations x_n, is defined by:

$$p(\boldsymbol{X}) = \prod_{n=1}^{N} p(\boldsymbol{X_n}) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \, N(\boldsymbol{X_n}|\mu_k, \Sigma_k) \qquad (17)$$

Following the same method applied to the original Gaussian density function, applying the log on either side of the equation (17):

$$\ln p(\boldsymbol{X}) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(\boldsymbol{X_n}|\mu_k, \Sigma_k) \qquad (18)$$

Online: ISSN 2645-2960; Print ISSN: 2141-3959

In order to find the optimal parameters for the Gaussian Mixture, the equation is derived from the parameters given in Zivkovic, (2014) results to equation (18).

3. Review of Related Literatures

Several researches have been carried out in the area of surveillance and motion detection. Authors in Ashish, Choubaey, Michin, and Funkar, (2019) developed a robust and efficient background subtraction algorithm that is capable enough to manage with local illumination changes, as well as, global illumination changes which are linked to a final system called Human Motion Detection System. Ashish *et al.,* (2019) proposed various techniques to effectively detect motion and maintain proper documentation of one or more objects moving and causing motion. This system is proposed to focus on fixed description area. Therefore, the background of such area is assumed to be non-moving and considerations of sudden change in lightning is ignored. The developed algorithm is successful and robust detection of human motion. However, the drawback of this system is that the model of human motion contains a static representation of what a human figure would look like, built over a period of time. This inherently cannot contain the information about the periodic nature of human gait. Thus, any moving object would trigger false alarm as there is no smart or secure control for this system.

Similarly, Smart Surveillance Monitoring System developed by Maya and Prasannaijit, (2018) deals with the development of a smart surveillance system using PIR sensor and Raspberry Pi for mobile devices. The device uses a 3G Dongle to transmit the captured information to the user's mobile smart phone. The system uses the PIR sensor to detect motion and initiate the camera to stream life video and records it for future playback. The drawback on this developed system is that the user must be online as to receive the alert as at when the incident occurs. That is, there is no means to tackle the issue if the user's device is not actively online and also the video playback is not well processed and as such the intruder's identity might not be well known and as such also deceive the user to activate the alarm system.

Sureindar and Saravanan, (2015) implemented a motion detection system on Raspberry pi using Lucas-kanade algorithm, which is an optical flow method to find displaced object from an image. The algorithm works by comparing two successive image frame and it forecast the direction of the displaced object. The main advantage of this algorithm is that it does not need to scan the next frame image for matching the pixel or neighboring pixel of the image and the process can also be achieved through optical flow vector and the vector will be similar to small neighborhood surrounding pixel. This system is limited with the fact that area of surveillance must not contain constantly moving objects and illumination changes affects the system in as much as the system is able to detect the direction of displaced objects in surveillance.

Furthermore, Shake and Borde, (2014) presented an integrated home monitoring system that assesses the implementation of a cost-effective alert system based on the detection of small movements. They worked at setting up low prices, low power consumption, good resource utilization and efficient monitoring system using a variety of different sensors. Their system allows real-time monitoring of home activities from anywhere and using a microcontroller that is now considered a limited resource and an open source solution to Single Board Computer. This system would yield false detection in cases that movements is not caused by an actual intruder but rather something that is not even human, also the power of the microcontroller used in this case the Arduino performed poorly in image processing.

Online: ISSN 2645-2960; Print ISSN: 2141-3959

In addition, Jeevanand, (2014) worked on the design of a network video capturing system using Raspberry Pi. The proposed system works on video captures and distribution with networked systems, as well as alerting the administrator through an SMS alarm, at the customer's request. Their system was designed to work in real-time situations and based on Raspberry Pi Single Board Computer (SBC). Unlike other integrated systems, their real-time application provides a client video monitor using the Alert Module and the Single Board Computer (SBC) platform. This system adopts the basic background subtraction technique in motion detection and as such is susceptible to false detection due to the fact that it does not cater for illumination changes, and the inability to mask certain arears in video surveillance.

Also, Singhd, (2015) described the IP Camera Video Surveillance system using Raspberry Pi technology. The researchers aim to develop a system that captures real-time images and displays them in a browser using TCP / IP. The algorithm for face detection is being applied to the Raspberry Pi, which allows live video streaming along with detection of human faces. The study did not include surveillance reactions in cases of intruder detected in video surveillance and does not detect motion in cases where intruder mask their faces. Similarly, Sebastian, Peter, and Philips, (2013) proposed an Edge based approach for robust foreground detection, this technique was developed to tackle the fallback of the Background Subtraction algorithm. this article proposed a framework for detecting moving objects in real life video processing applications under various lighting conditions. It illustrates the robustness of a combination of edge detection and smoothening techniques. Due to the adaptation approach it adopts to tackle the illumination changes, it affects the overall performance of the system, also another minor drawback is the issue that the system has not yet fully light-incentive thresholds. Further exploration to automatically adapt the thresholds to the light changes is required.

Furthermore, Chauhan and Anuradha, (2015) designed and developed a real-time video surveillance system based on the Raspberry PI B + Board embedded web server. Their system has a low cost, good openness and is easy to carry and easily maintained and improved. Thus, the application system provides a better security solution. The system can be used to influence security in the industrial and residential areas, as well as, military formations, but has no means of detecting motion or intrusion as they did not employ any motion detection algorithm in the design.

In addition, Madhavi, Rutuja, Neelam, and Amol, (2014) in its review developed a security-based home security using android application. The researchers developed a mobile robot that is controlled from a distance by a human operator through a GUI (Graphical User Interface). The proposed application includes the simulation of the sensors to predict the actions of the robot. The drawback of this system is that it is not secured as it could trigger alarm on wrong detection by the robot because the alarm is triggered before the video footage is being sent to the user's mobile device through SMS and also the android application.

In the article presented by Murungi, (2009), a video surveillance system making use of IP based surveillance to monitor the environment round the clock i.e. 24 hours' daily was proposed. The system makes use of distributed cameras connected across a network using their IP, this research makes use of IOT function in its development, the drawback of this system is such that the system has little attention to the cost and also increased false detection rate of moving objects since they employed the basic background subtraction technique.

Authors in Jadhav, (2014) evaluated the use of various sensors, wireless module, microcontroller unit and fingerprint module to formulate and implement a cost-effective monitoring system. The authors adopted an ARM

Online: ISSN 2645-2960; Print ISSN: 2141-3959

kernel as the core processor of the system. The PIR sensor is used to detect motion in the viewing area, while the vibrating sensor was used to detect any vibration event such as a breakout sound. The intruder detection technique is proposed using the PIR sensor which detects the movement and triggers a system of alerting and sending short message service via a Global Service Messenger GSM module for a specified telephone number. Their work can be implemented on any database, it will be safer and more difficult to hack. Table 1 shows comparison of these related works in the subject matter.

Table 1: Comparison related works in Smart Surveillance Systems

S/N	**Authors Name (Year of publication)**	**Title**	**Technique applied in Motion Detection**	**Sensors used**	**Ability to Adapt**	**Ability to Mask**
1.	Ashish, *et al.* (2019)	Motion Detection Surveillance System using an improved background subtraction algorithm	Update-based Background Subtraction	No	Yes	No
2.	Harum *et al.,* (2018)	Smart Surveillance System using Background Subtraction Technique in IoT Application	Background Subtraction	No	No	No
3.	Sureindar, and Saravanan, (2015)	Motion Detection using optical Flow on Raspberry Pi	Optical Flow	No	No	No
4.	Shake, Padmashree and Borde, Sumedha (2014)	Embedded surveillance system using PIR sensor	Sensor triggered	Yes	Yes	Yes
5.	Jeevanand,(2014)	Real Time Embedded Network Video Capture and SMS Alerting system	Background Subtraction	No	No	No
6.	Maya, *et al.,* (2018)	Smart Monitoring System using Raspberry Pi and PIR sensor	Sensor Triggered	Yes	Yes	No
7.	Singhd, (2015)	IP Camera Video Surveillance using Raspberry Pi.,	Optical Flow Background Subtraction with facial recognition	No	No	No
8.	Sebastian *et al.,* (2013)	An Edge-based Approach for Robust Foreground Detection	Edge-detection based approach	No	No	Yes
9.	Chauhan, & Anuradha. (2015)	Design and Develop Real Time Video Surveillance System Based on Embedded Web Server Raspberry Pi B+ Board	Sensor triggered	Yes	Yes	No

Online: ISSN 2645-2960; Print ISSN: 2141-3959

4. Analysis of Existing Systems

On close observation of the reviewed literatures, the existing systems have some or all of the following components integrated:

i. Controller (Arduino or Raspberry Pi)
ii. Camera Module
iii. GSM module
iv. Alarm system (Buzzer)

The procedure in using the established surveillance system is illustrated in Figures 5 and 6.

The procedure for the existing system as illustrated in Figure 6 can be summarised as follows:

i. The video is converted to frames of images and taken in as input
ii. The reference image is extracted from the Sequence of images
iii. The new image is subtracted from the reference image
iv. A condition is stated for each of the pixel variant if it is greater than the threshold
    a. If yes, the movement of the frame is tracked and motion/foreground object is detected
    b. Else if No, then no motion detected end.

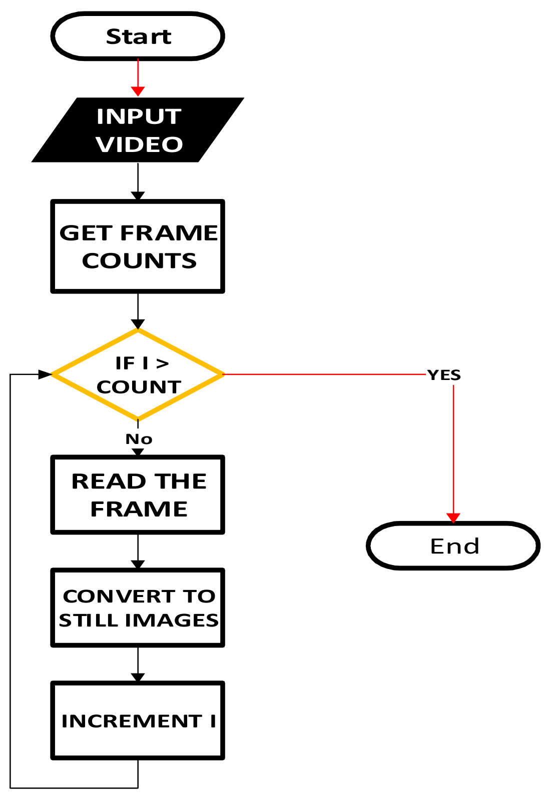

Figure 5: Flowchart on Image Frame Acquisition from video (Source: Ashish, Choubaey, Michin, & Funkar, 2019).

Consequently, the critical analysis of the review of related works in the section three revealed the ability of existing algorithms to control and access surveillance remotely but fails to tackle adaptation to illumination changes. Most of the reviewed papers presented background subtraction technique, which requires static background. The background subtraction technique as such does not work for varying background and foreground. In this paper we improve on the limitations presented by Ashish & Choubaey,(2013); Maya and Prasannaijit (2018); Sebastian *et al.*, (2013); Murungi, (2009) and Ashish *et al.,* (2019) by applying the proposed adaptive Gaussian Mixture model in improving motion detection by making use of the network module for the remote communication and additional image processing technique to tackle for adaptation to both foreground and background changes.

5. Proposed Design and Methodology

The embedded system consists of Raspberry Pi3, which is the central controller for all the image processing and logical programming for the whole system comprising Raspberry Pi3 8MP camera module with auto focus lens and GSM module and buzzer (Alarm notification).

The system comprises of four basic components

i. Raspberry Pi 3
ii. Raspberry Pi 3 8MP Camera module
iii. GSM module
iv. MicroSD card

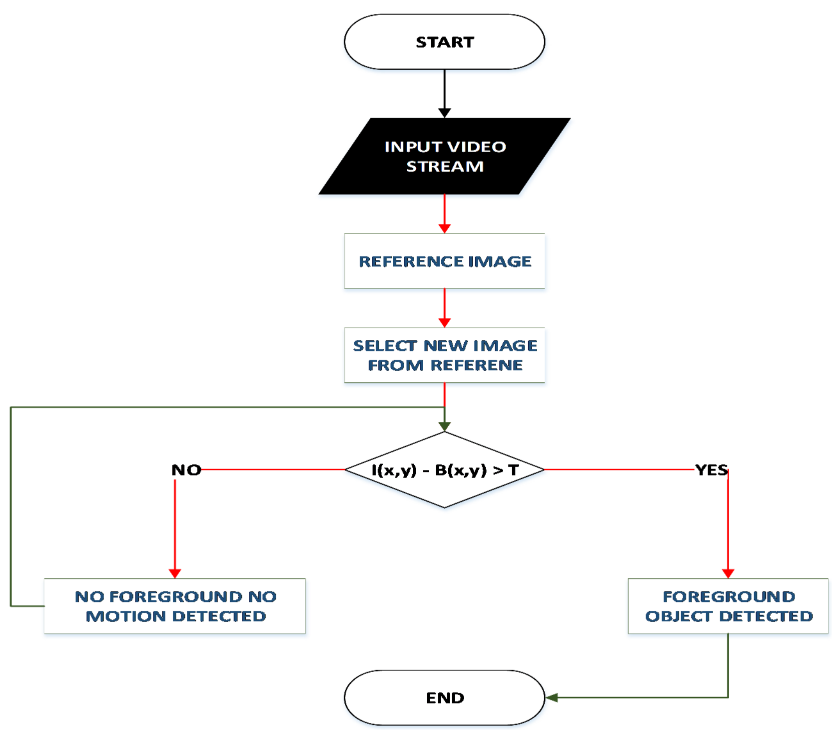

Figure 6: Flowchart of motion detection method (Source: Ashish & Choubaey, 2013)

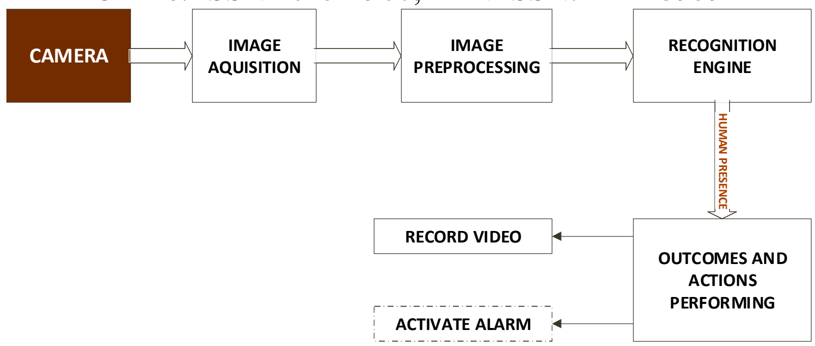

Figure 7: Block diagram of an existing system (Source: Ashish & Choubaey, 2013)

Raspberry Pi board is a core module for image/video capture, processing the acquired image/frames, performing background subtraction algorithm on the images acquire to detect objects. The Board has ARM cortex A53 clocked at 1.2GHz, 4000MHz Video Core IV multimedia GPU, 1Gb memory, power supply, HDMI, USB ports and other features (Rundle, 2016) as shown in Figure 8. Likewise, Figure 9 presents the bird-eye view for flow diagram of the proposed surveillance system. The flow diagram shows the sequential ordering of core system processes and their linear interactions. The camera would be responsible for motion detection by applying the adaptive Gaussian mixture technique to improve the background subtraction algorithm.

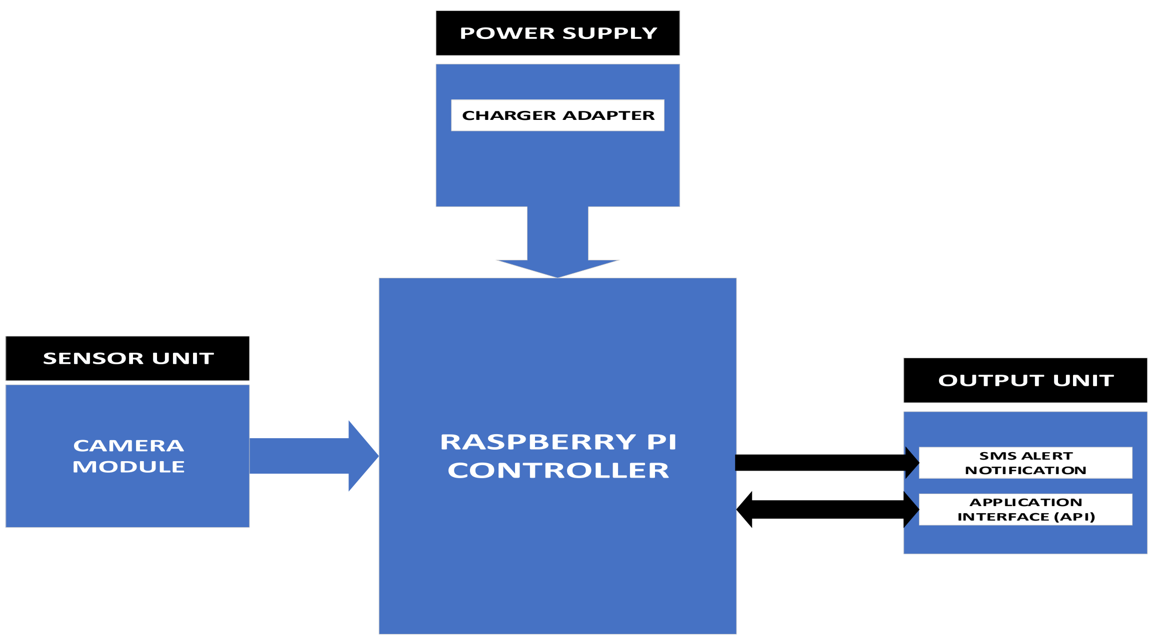

Figure 8: Block Diagram of the Proposed Secure Surveillance System

In order to implement the proposed system, the electronic components to be used are; Raspberry pi 3, Pi-Camera module, Switch button, power supply. The camera, the push button and the power supply all act inputs to the microcontroller (Raspberry pi3) while the alert system acts as the output for the system. The API (application interface) acts as both input and output in the surveillance system. Once images/frames are acquired from the camera and fed into the Raspberry pi3 controller, the image is being processed using the enhancement techniques by applying the adaptive Gaussian Mixture model to model the background for background subtraction as shown in Figure 10. The images that undergo background subtraction are stored within the programmed Raspberry pi3.

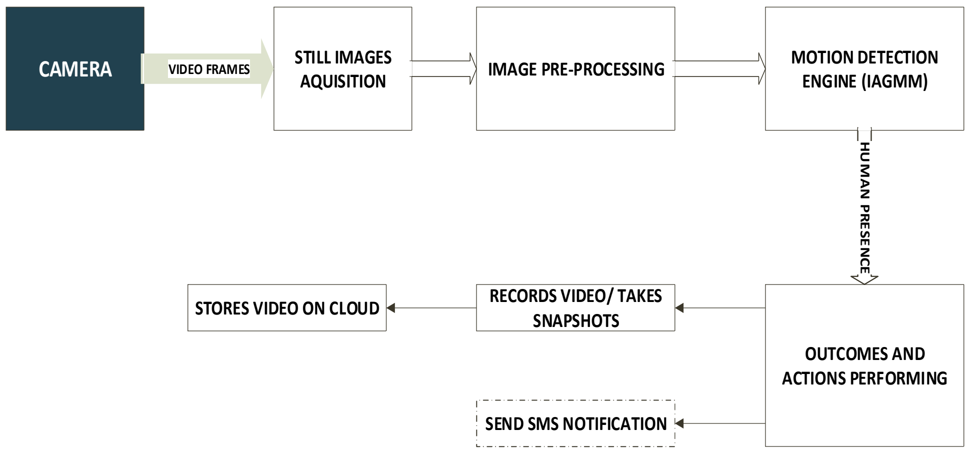

Figure 9: Flow Diagram of the Proposed Smart Surveillance System

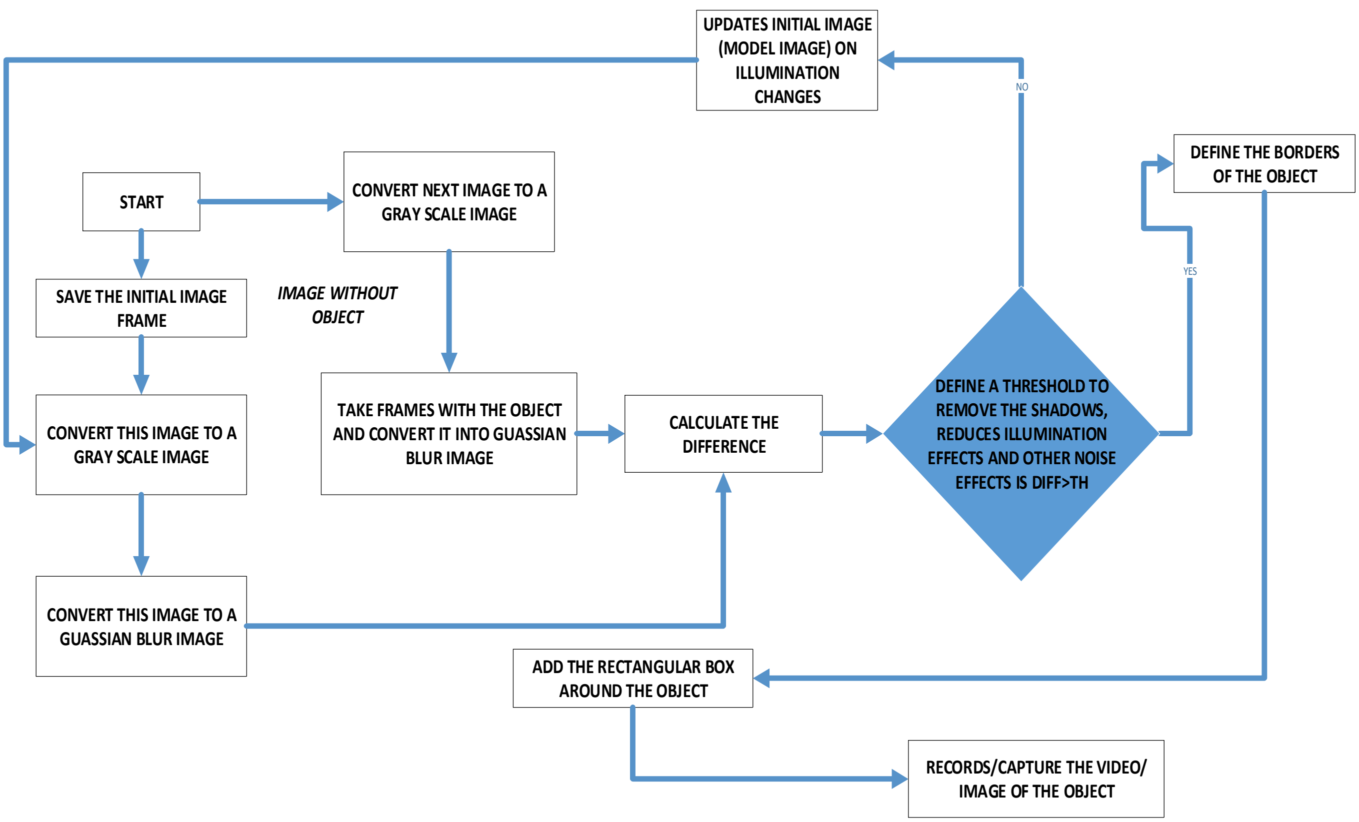

Figure 10: Flow diagram of the developed Enhanced Background Subtraction Technique (AGMT)

6. Conclusion and Future Works

Security is a top priority for individual and corporate bodies from both local and global perspectives. Interestingly, the findings from the various articles reviewed show that background subtraction technique is promising in terms of performance and efficiency, however, there exist several areas where this approach needs to be further improved. These include inability to mask certain areas and accounting for illumination changes. Thus, the need to enhance the background subtraction techniques. This paper presented the systematic reviews of background subtraction techniques for smart surveillance system and the design phase of a proposed smart surveillance system to address the identified shortcomings of background subtraction technique identified from literatures. Implementing the proposed system will boost detection of motion objects in and around residential and commercial areas. Thus, this study intends to develop the proposed surveillance system and evaluate its performance in our future research on the subject matter.

REFERENCES

Ackerman, S., & Thielman, S. (2016). US intelligence chief: we might use the internet of things to spy on you. US: The Guardian.

Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., & Moriarty, T. (2002). A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data. IEEE Transactions on Medical Imaging, 193-199. doi:10.1109/42.996338

Alireza, K., Kashani, A., Milani, N., Akhlaghi, P., & Khezri, K. (2008). Robust Color Classification Using Fuzzy Reasoning and Genetic Algorithms in RoboCup Soccer Leagues. Robocup. Lecture Notes in Computer Science, 548–555. doi:10.1007/978-3-540-68847-1_59

Ashish, k., & Choubaey, A. (2013). Motion Detection Surveillance System using background subtraction algorithm. Advance Research in Computer Science and management studies.

Ashish, k., Choubaey, A., Michin, A., & Funkar, I. (2019). Motion Detection Surveillance System using an improved background subtraction algorithm. Advance Research in Computer Science and management studies.

Azernikov, S. (2008). Sweeping solids on manifolds. In Symposium on Solid and Physical Modeling, 249-255.

Banerjee, T. (2014). Day or Night Activity Recognition From Video Using Fuzzy Clustering Techniques. IEEE Transactions on Fuzzy Systems, 483–493. doi:10.1109/TFUZZ.2013.2260756

Bezdek, C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press.

Canny, J. (1986). A computational approach to edge detection. Pattern Analysis and Machine Intelligence, 679-698.

Chauhan, M. F., & Anuradha, G. (2015). Design and Develop Real Time Video Surveillance System Based on Embedded Web Server Raspberry Pi B+ Board. International Journal of advanced Engiineering and Research Development (ijaerd), 1-4.

Chen, S. S. (2008). Moving object segmentation using improved running Gaussian average background model. Proceedings of the International conference on Digital Image .

Cheng, L., Gong, M., Schuurmans, D., & Terry. (2011). Real Time Discriminative Background. IEEE Transactions on Image Processing, 1401-1414.

Davis, E. (2005). Signal Processing and its application (3rd ed.). Morgan Kaufmann. doi:10.1016/B978-0-12-206093-9.50039-3

Fujika, Plakninly, & James. (2017). The Surveillance Studies Centre. Queen's University: A multi-disciplinary research centre at Queen's University.

Harum, N. b., Ali, M. F., Zakaria, N. A., & Anawar, S. (2018). Smart Surveillance System using Background Subtraction Technique in IoT Application. Centre for Advanced Computing (C-ACT), Faculty of Information and Communication Technology, IV(12).

Harville, M. (2002). A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models. In Proceedings of ECCV.

He, Z., Liu, Y., Yu, H., & Ye, X. (2008). Optimized algorithm for traffic information collection in an embedded system. proceedings of the IEEE Congress on Image and Signal Processing.

Hornberg, A. (2017). Handbook of Machine and Computer Vision. Wiley-VCH Verlag GmbH & Co. KGaA.

Jadhav. (2014). Real Time Embedded Network Video Capture And SMS Alerting system,.

Jeevanand, D. (2014). Real Time Embedded Network Video Capture And SMS Alerting system.

Kato, J., Joga, S., Rittscher, J., & Blake, A. (2002). An HMM-based segmentation method for traffic monitoring movies. IEEE Trans. on PAMI, 1291–1296.

Lyon, D. (2001). Surveillance Society. Buckingham: Open University Press.

Lyon, D. (2007). Surveillance Studies. Cambridge: Polity Press.

Madhavi, S., Rutuja, S., Neelam, T., & Amol, B. (2014). Security Based Home SUrveillance System Using android application. International Journal of Research in Engineering and Technology.

Mai, F., Hung, Y., Zhong, H., & Sze, W. (2008). A hierarchical approach for fast and robust ellipse extraction. Pattern Recognition, 2512-2524.

Marx, G. T. (2015). Surveillance Studies. International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 733-741.

Maya, N., & Prasannaijit, D. (2018). Smart Monitoring System using Raspberry Pi and PIR sensor. Indian Jornal of Research, 7(6).

Mori, G., Ren, X., Efros, A., & Malik, J. (2004). Recovering Human Body Configurations: Combining Segmentation and Recognition. IEEE Computer Vision and Pattern Recognition.

Murungi, K. (2009). Video Surveillance System.

Park, J., Tabb, A., & Kak, A. C. (2006). Hierarchical data structure for real-time background subtraction. The proceedongs of the International Conference on Image Processing (ICIP ’06), 1849–1852.

Prince, S. (2012). Computer Vision: Models, Learning, and Inference (1st ed.). New York, NY, USA : Cambridge University Press.

Ren, X., & Malik, J. (2003). Learning a classification model for segmentation. Proc. 9th Int. Conf. Computer Vision, volume 1, 10-17.

Rundle, M. (2016, Feburary 29). Raspberry Pi 3 the first with built-in wi-fi. Retrieved May 20, 2019, from http://www.wired.co.uk/article/raspberry-pi-three-wifi-bluetooth-release-price-cost

Schaeffel, F., & Hornberg, A. (2017). Handbook of machine and Computer vision. doi:10.1002/9783527413409.ch1.

Sebastian, G., Peter, V. H., & Philips, W. (2013). An Edge-based Appraoch for Robust Foreground Detection. Gent, Belgium: Ghent University TELIN-IPI-IBBT, Sint Pietersnieuwstraat.

Shake, P. A., & Borde, S. S. (2014). Embedded surveillance system using PIR sensor (Vol. II).

Singh, A., Sawan, S., Hanmandlu, M., Madasu, V. K., & B. (2009). An abandoned object detection system based on dual background segmentation. Proceedings of the 6th IEEE nternational Conference on Advanced Video and Signal Based Surveillance (AVSS '09), 352–357.

Singhd, S. (2015). IP Camera Video Surveillance using Raspberry Pi.,.

Stauffer, C., & Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In Proceedings CVPR, 246–252.

Sureindar, C., & Saravanan, D. P. (2015, November). Motion Detection using optical Flow on Raspberry Pi. International Research Journal of Engineering and Technology (IRJET), II(8).

Yang, Y. (2009). Image segmentation based on fuzzy clustering with neighborhood information. Optica Applicata.

Zareiforoush, H., Minaei, S., Alizadeh, M., & Banakar, A. (2015). Potential applications of computer vision in quality inspection of rice: a review. Food Eng. Rev. 7, III, 321-345. doi:10.1007/s12393-014-9101-z

Zheng, Y., & Fan, L. (2010). Moving object detection based on running average background and temporal difference. proceedings of the IEEE International Conference on Intelligent Systems and Knowledge Engineering (ISKE '10), 270–272.

Zivkovic, Z. (2014). Improved Adaptive Gaussian Mixture Model for Background Subtraction. Netherlands: Intelligent and Autonomous Systems Group.