Adaptive Background Modeling for Dynamics Background

N.A. Zainuddin, Y.M. Mustafah, A. A. Shafie, A.W.Azman, M.A. Rashidan, and N.N.A. Aziz

Department of Mechatronics, Kulliyyah of Engineering, International Islamic University Malaysia

Abstract
An increasing number of CCTV have been deployed in public and crime-prone areas as demand for automatic monitoring system is increasing to counterbalance the limitation of human monitoring. To have a good monitoring system in such places, a good background model is needed in order to reduce amount of the video processing needed for tracking, classification, counting and etc. This paper proposes an adaptive background modeling that is able to model a scene under review at real-time. The proposed modeling system is also expected to be able to handle dynamic backgrounds and common problems in detection methods. A novel patch-based background reconstruction based on highest frequency of occurrences assumption and past pixel observation is proposed. Contrast adjusting method is used to reduce the problem of incorrectly classified foreground which is shadow problem. The proposed algorithm is focused to be tested and analytically compared with the dynamic background at the indoor and outdoor environment. The main challenges of background subtraction such as illumination changes, geometrical changes, stationary moving object problem and high speed object problem are taken care of and extensively discussed in this paper. The experimental results show that the algorithm is able to reconstruct a background model and produce accurate and precise foreground that can be used for other processing stages.
Keywords Adaptive background modeling; Background subtraction; Surveillance; Foreground segmentation; Dynamic background.

1 Introduction

There are two basic elements in video analysis that need to be identified, which are background and foreground region. The background which is also known as the reference frame usually consists of non-interested objects, either static or dynamic. Foreground image on the other hand, is the objects of interest that need to be identified, detected or further analyzed. The shape of a foreground region unnecessarily be a rectangle. It can be arbitrary in shape and the background is in the complementary shape of the foreground region. The most common technique used by many researchers in handling this problem is by using background subtraction[1,2]. There is a lot of study in background subtraction, but the problem is still yet to be properly solved. The problem of *ghost*, stationary moving object and repetitive movement of complex background scenes have been attempted to be solved, however, the result is still unsatisfactory. In the early days, many researches have focused on improving the general background subtraction process such as temporal differencing, double temporal differencing and many more. The recent trend shows that researchers tend to work on the specific stage in the background subtraction algorithm, which is background modeling and background reconstruction. The reconstruction of the background model is essential in order to correctly identify the foreground. A good background model will benefit the analysis of the foreground and produce accurate and effective background subtraction. The simplest method is setting the image without any moving object as the background[3]. However, this assumption is too idealistic, and not possible to solve many problems especially in outdoor environments, such as on a crowded street with vehicles and pedestrians or an outdoor environment like in a street will experience various levels of illumination at different times of the day[4]. Plus, outdoor environment has a tendency of experiencing adverse weather condition like fog, rain or strong wind that could modify the reference image[5]. In these cases, the background must be adaptively refreshed and updated. Therefore, in this paper, an adaptive background modeling technique is presented. There are five sections in this paper. In section two, previous works related to the research are described. Meanwhile, in section three, the proposed algorithm is presented and discussed extensively. In section four, the experimental results at various scenes are presented with qualitative and quantitative results from various situations are illustrated. Finally, the conclusion is drawn in section five.

2 Related Works

Recently, a lot of works are concentrating on adaptive background reconstruction[6-9]. Several notable methods were introduced including temporal smoothing, pixel intensity classification, running Gaussian average, Gaussian mixture model, hidden Markov model and kernel density estimation.

Gaussian mixture model (GMM) is a common method used by researchers in the foreground detection. The main idea of the Gaussian mixture model is by having N number of Gaussian distribution function, in order to reconstruct the background pixel. The implementation of GMM has been proposed as the adaptive background reconstruction technique, especially in surveillance system since late 1990s by *Stauffer et al.*[10]. As it is still in the early stage, a lot of unwanted small blobs are scattered all over the frame and the main blobs aren't completely detected as foreground. Since then, many of the researchers have continued to improve the proposed algorithm[11-13]. *Mukherjee et al.* have incorporated the Horpreset colour model in order to make the algorithm to be able to detect shadows[11]. *Mukherjee et al.* work focuses more on the shadow handling, the experimental results show that the algorithm is tested for a non-complex indoor scene with one foreground handling. Literature[12,13] managed to reconstruct the background with shadow consideration better with a more complex scene than *Mukherjee et al.*'s work. They use the train station dataset in order to have multiple object handling. They have implemented the use of window-based decision rules for the shadow and the background model is reconstructed by GMM. To sum up, many implementations of GMM are focusing on the indoor scene for the surveillance system.

Elgammal et al. claimed that GMM is ideal for indoor scenes only[14]. Thus, they introduced the use of kernel density estimation (KDE) for background reconstruction. In KDE method, the background and foreground pixel are model of probability density function (pdf) and the pdf is estimated by a *kernel* function or also known as *window function*. However, the major drawback of this method is the expensive computational cost. Later *Gao et al.* improved the method by introducing the Marr wavelet equation in estimating the pdf[5]. Marr wavelet is a second derivative of the Gaussian smooth function. They combine the Gaussian and Kernel method together. However, from the experimental results, the background model is not reconstructed correctly as some of the background pixels are counted as a target; thus in some cases, the target is not detected correctly. *Lee et al.* managed to reduce the need of storage[15]. They initialized the first frame and update it at every frame by setting the learning rate. For dynamic background cases, they used threshold method. The first frame was set as the background model and later, updated by learning method. This method is only ideal for the video sequence at has no target or foreground the beginning of the video sequence. For the crowd and complex scene, it is quite impossible to get such video sequence.

The other method that commonly used by researchers in reconstructing the background is temporal smoothing[16]. The main idea of temporal smoothing is combining the stored image with the new image on pixel-based. Pixels on the smoothed image will be replaced by a part from previous value combined with a new value from new image at the same position. *Ridder et al.*[17] have improved the method by modeling each pixel with Kalman Filter. They managed to make the system more robust to cope with the illumination changes, however, the algorithm update the background slowly. Later, in recent year, *Hung et al.* improved the method by combining with the median filtering[6]. By doing so, they managed to reduce the computing frequency of median operations. However, they focus on the computational performances of the algorithm with no real life situation is considered in their experiment. Motivated by [6], *Asif et al.* use the temporal smoothing and median filtering for modeling the background[7]. As their research focus more on the human gait, the background and foreground are non-complex. Based on their experimental result, the temporal smoothing seems to be ideal for the non-complex scene. Thus, the implementation of temporal smoothing algorithm is ideal for surveillance system that involves non-complex scene (indoor surveillance).

In early of 2000s, *Hou et al.* have introduced the pixel intensity classification (PIC) method as the background reconstruction technique[18]. In this approach, at every frame, the difference of pixel intensity is calculated. Then, the classification is made based on the calculated difference. The background model is assumed to be the highest frequency in the intensity value. The simulation results of *Hou et al.* Literature[18] used the outdoor dataset with parking lot environment. The works on the same method is continued by *Xiao et al.* in 2006 and 2008[2,8]. *Xiao et al.* managed to handle more adverse weather by adapting the raining situation, however, their experimental results shown that the adaption of the algorithm to the rain is still at the early stage. Next, *Cao et al.* have employed the improved version of PIC in reconstructing the background model for light flow traffic movement video sequence[19]. They managed to compare their experimental results with GMM and Time averaging algorithm. The results however are only ideal for slow moving to medium moving foreground detection. Some of the foreground in high speed foreground is not detected and small blobs appeared as a result from the *ghost* problem also known as the blending of high speed object movement in the video sequence.

After analyzing these methods, several assumptions were made. In non crowded scenes, the background pixel would be the maximum frequency in the image sequence. For crowded condition, most of the pixels in the image frame are expected to be modified. From the assumptions, we developed a background reconstruction algorithm based on pixel intensity classification and mode filtering. However, the difference in the inter-frame pixel intensity value is not calculated in the first step, the mode filtering is done first in order to model the background.

3 Methodology

In order to produce a good background model, several techniques are deployed through several stages with pre-determined assumptions. Fig.1 shows the overall process of proposed background modeling algorithm. An elaborated discussions on the proposed algorithm is presented in this section.

3.1 Pre-processing

The video will be converted into uniform size of image sequences. Next, the image sequence will be gray-scaled and undergone median filtering process in order to get rid of noises, especially from the camera pixel noise and impulse noise.

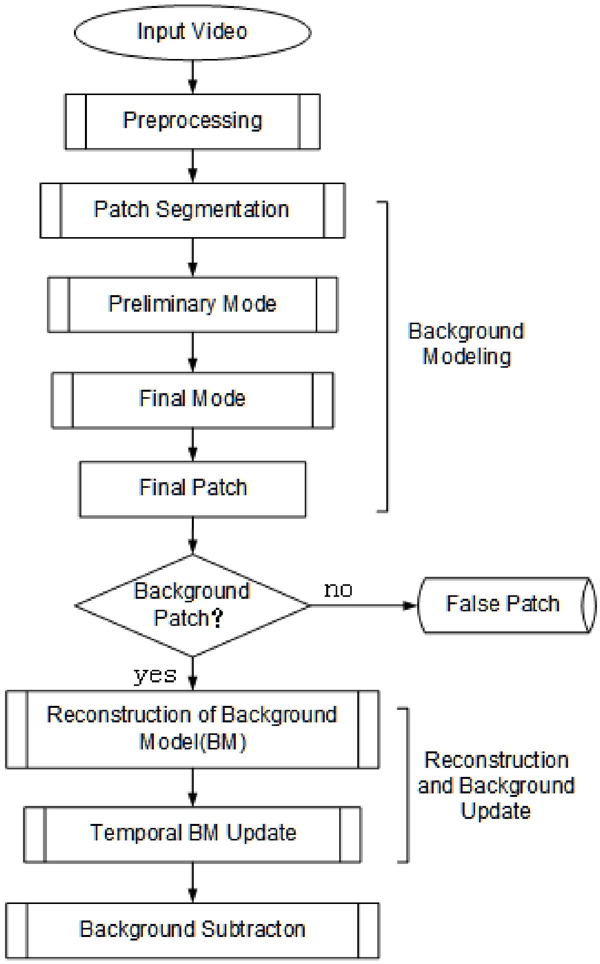

Fig. 1 Process of proposed background modeling algorithm

3.2 Background Model

In order to model the background, an assumption of background-foreground pixels needs to be made. Thus, based on the assumption that the background pixel is the most frequent appeared in the entire video sequences, the proposed algorithm is as follows:

Step 1 : Calculate the mean of each patch

Let $f_1, f_2, f_3, \cdots, f_n$ represent the frames from the same video sequence. First, each of the frames must be segmented into $m \times n$ patches, where m and n can be the same number. Assume N patches are obtained, where we let patches be marked as $p_1, p_2, p_3, \cdots, p_N$. Experiments are conducted to find the optimum size of the patches. The result of the experiments shows that the size of the patches is not as sensitive as the sampling frame value. It does affect the result, but in small percentage error. Then, at each frame,$f_1, f_2, f_3, \cdots, f_n$ the intensity value of pixel (x,y) in the area of specified patch, p_i are marked as $i_1, i_2, i_3, \ldots, i_n$, where $i = 1, 2, 3, \ldots, n$ and total number of pixels in one patch can be written as:

$$q = m \times n \tag{1}$$

Then, mean $m_{p,f}$ at each frame for the specified patch is calculated as in the formula below:

$$m_{p,f} = \frac{\sum_{j=1}^{q} i_j(x, y)}{q} \tag{2}$$

Step 2 : Calculate the preliminary mode

In order to calculate the mode, the calculated mean is grouped into matrices called $mean\_matrix_{p,f}$ according to the specified sampling time. Assume that the sampling time is denoted as T. Thus, the preliminary mode at that time interval for patch 1 is:

$$n_{modepre} = max(m_{1,1}, m_{1,2}, \cdots, m_{1,T-1}) \tag{3}$$

This formula is repetitively used for all other patches.

Step 3 : Check the preliminary mode

There are two possibilities of the mean matrices. First, there is repetitive mean value and the second one is there is no repetitive mean value. For the former case, there is no issue as it will follow equation 3 correctly. However, for the latter case, the result of the mode is not correctly calculated. The mode algorithm will set the minimum mean value as mode if there is all frequencies of the matrix elements are 1. To avoid this, the checking algorithm is set. Assume that the result is the final mode and marked as,$n_{modefinal}$ and a threshold value, ε is introduced. Meanwhile, $sub_{p,f}$ matrices are introduced to store the result of subtraction. If the number of zero in the $sub_{p,f}$ matrix equal to or greater than a threshold number, ε, the final mode, $n_{modefinal}$ is set to the calculated preliminary mode, $n_{modepre}$ indicating there is repetitive mean value in the $mean\_matrix_{p,f}$. Likewise, if the number of zero in the $sub_{p,f}$ matrix less than a threshold number, ε, the final mode, $n_{modefinal}$ is set to the previous frame values, as there is repetitive mean value in the $mean\_matrix_{p,f}$.

$$sub_{p,f} = |mean\_matrix_{p,f} - n_{modepre_{p,f}}| \tag{4}$$

$$n_{modefinal_{p,f}} = \begin{cases} n_{modepre_{p,f}} & n(sub) = 0 \geq \varepsilon \\ n_{modepre_{p,f-1}} & n(sub) = 0 < \varepsilon \end{cases} \tag{5}$$

Step 4 : Construction of background model and adaptive update

To build the background model, the calculated final mode values are utilized. It is important to note that, in the earlier stage, we have built storage to store the patches in RGB image form matrices. Thus, the final stage to build the background model is by pointing the right RGB image of patch to the right patch according to the final mode value.

3.3 Background Subtraction and Binarization

In binarization process, the images will be transformed into black and white images. The binarization algorithm used in this research is motivated by global-used binarization method, Otsus method, in which threshold algorithm is deployed. This method is computationally inexpensive thus ideal for real time applications. In this method, we need to set for an initial threshold, T_i. We estimate average of the minimum and maximum pixel value of the image, T_i, where mathematically it is calculated as:

$$T_i = \frac{max(i) + min(i)}{2} \tag{6}$$

Then, the whole pixels are segmented based on the T_i value. For the intensity values that greater than or equal to the T_i value, it is grouped into G_1. Otherwise, the intensity values will be grouped to G_2 and the average intensity values of each group are calculated and noted as ave_1 and ave_2 respectively, as shown in equation 7,8,9 and 10 respectively.

$$Group = \begin{cases} G_1 & i_j \geq T_i \\ G_2 & i_j < T_i \end{cases} \tag{7}$$

$$ave_1 = \frac{\sum i_{G_1}}{n(i_{G_1})} \tag{8}$$

$$ave_2 = \frac{\sum i_{G_2}}{n(i_{G_2})} \tag{9}$$

Based on the ave_1 and ave_2 value, new threshold value is computed:

$$T_{final} = \frac{ave_1 + ave_2}{2} \tag{10}$$

3.4 Shadow Removal

Shadow is the incorrectly classified foreground pixels that need to be eliminated. Foreground mask without shadow removal will adversely affect further processing by providing the false information such as inaccurate blob size, centroid and etc. Thus, the efficiency of the method will be affected. There are many proposed algorithms for shadow removal, however, as we already have a good background model based on the proposed algorithm, we just need a simple shadow removal. Thus, the selection of algorithms of the shadow removing is focused on the computational aspects. One of the simplest methods for shadow removal is by adjusting the contrast. To simplify that we just adjust the luminance value of the pixel by multiplying the current luminance value, $y_{current}$ with a predefined contrast factor, C. Thus, the new luminance value is:

$$Y_{new}(x,y) = y_{current}(x,y) \times C \tag{11}$$

3.5 Morphological

In order to filter the small unwanted pixel in the binary images of the foreground mask, the operation of mathematical morphological is used. There four are basic morphological operations, which are erosion, dilation, opening and closing. In this research, we used the closing morphological operator to close the gap of highly-deformed of interested object shape. Closing morphological operator mathematical definition is basically the morphological dilation followed by erosion operation. Thus, the equations are:

$$H \bullet I = (H \oplus I) \ominus H \tag{12}$$

where I is the binary image of the foreground and H is the structuring element. Holes in the foreground those are smaller than H will be filled. Thus, it will reduce the deformation on interested shapes.

4 Experimental Results

In this section, evaluation of the methodology is presented in order to compare the proposed algorithm with several existing methods. Both qualitative and quantitative perspectives are discussed. In order to prove the robustness of the proposed algorithm, several video sequences with various scenes are tested in which including the complex dynamic background at the indoor and outdoor environment.

4.1 Comparison Methods

Several works on background modeling and object detection techniques are chosen to be compared in terms of performance with our proposed method. The methods are stochastic approximation[20], Bayesian learning-based[21], adaptive Gaussian mixture model[22], Gaussian mixture model[23], and Pfinder[24]. All the methods are tested using Matlab software and performed on Windows 7 on Intel ® core^{TM} i3-2330M CPU 2.2GHz processor.

4.2 Performance Measurements

To quantitatively evaluate the performance, well-known parameters in gold standard test measurements[25,26] are used. In this method, there are four important parameters that need to be defined which are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). In this research, the true positive is defined as the number of pixels that correctly identified as foreground, conversely false positive is number of background pixels that incorrectly identified as foreground. Meanwhile, true negative is defined as the number of pixels that correctly identified as background. On the contrary, false negative refers to number of foreground pixels that incorrectly identified as background[27]. From these four parameters, the performance measures are defined as follows:

$$Recall(TPR) = \frac{TP}{TP + FN} \tag{13}$$

$$Precision(PPV) = \frac{TP}{TP + FP} \tag{14}$$

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{16}$$

$$Specificity(TNR) = \frac{TP}{TP + FN} \tag{17}$$

$$Negativepredictivevalue(NPV) = \frac{TN}{TN + FN} \tag{18}$$

$$FallOut(FPR) = \frac{FP}{TN + FP} \tag{19}$$

$$Falsenegativerate(FNR) = 1 - TPR \tag{20}$$

$$FalseDiscoveryRate = 1 - PPV \tag{21}$$

4.3 Results

Fig.2 shows the results of the studied methods with our proposed methods in *MovingCurtain, WaveBeach, CampusRoad* and *Fountain* video sequences1. Each row shows the foreground mask generated by the corresponding methods for each video sequence respectively. Row 2, row 4, row 6 and row 8 of Fig.2 show the overlapping results of the foreground generated after background subtraction and ground truth. The overlapping results are used to quantitatively measure the performances of each method. The true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are defined based on the overlapped pixel colours. True positive denoted by white pixels and black pixels are true negative. Meanwhile, magenta pixels are for false positive and green pixels indicate false negative. On the other hand, the first column shows the selected frames from the respective video sequences and the second column shows the ground truth information of that frame. The other columns are the foreground mask results from the studied methods and our proposed results.

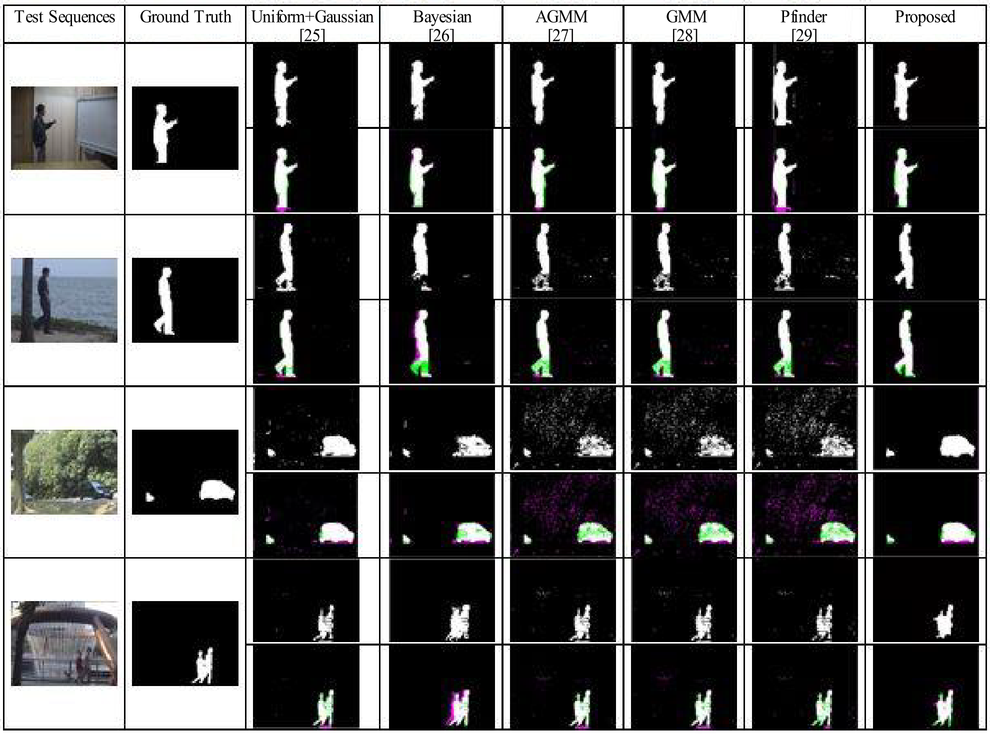

Fig. 2 Experimental results on complex scenes of dynamic background
Frame number *for MovingCurtain, WaveBeach, CampusRoad and Fountain* are 2774,1499, 2348 and 1196 respectively

From Fig.2, we can deduce that in the high variability scenes like *MovingCurtain* and video sequences, all algorithms handle the continuous moving curtain quite well, except for Pfinder[24] as several of the moving curtain pixels are falsely detected as foreground and shadow suppression also failed to be handled by Pfinder[24] and Uniform-Gaussian[20] that causing the additional white pixel formation at the bottom of human foreground for both foreground masks. In the *WaveBeach* video sequences, the variability of the background scene increases as the wave at the beach are continuously moving at a higher frequency than the moving curtain, plus as it is in an outdoor environment, the illumination condition is tested with the addition of stationary moving object condition. This is because the foreground (human) in this sequence stays for quite a long time before moving, thus, the moving object (human) tends to be falsely classified as background. In this video sequence, we can see that most of the algorithms could not handle the continuously moving wave as several pixels are falsely classified as foreground except for our proposed algorithm and uniform-Gaussian[20] that able to give a clean foreground mask. For the stationary moving object cases, most of the algorithms lost the information of bottom part of the human (foreground) however, the upper part of the foreground are perfectly detected.

In the *CampusRoad* video sequences, the problem of similar pixel colours and intense noise are evaluated. Out of five methods, two methods are able to separate the noise from the negative effect of the waving trees which are our proposed algorithm and the Bayesian[21] method. As the dark-coloured car is quite similar to the colour pixels of background, the shape of the car in the foreground mask is deformed except for our proposed algorithm and uniform-Gaussian[20]. Meanwhile, in the *Fountain* video sequences where the dynamic background (running water from the fountain) is more than 70 percent of the frame size and the pedestrian keep moving in the entire video sequences requires the algorithm to be able to update the background model in a short time. The problem of *'ghost'* detection or slow update can be detected in Bayesian approach foreground mask as the foreground mask tend to be bigger in size than the actual size. Meanwhile, the repetitive movements from the fountain fail to adapt by adaptive Gaussian mixture model (AGMM), GMM and Pfinder.

To quantitatively measure the performance of the studied methods and proposed method,the performance measurements discussed in Section 4.2 are used and tabulated in Table 1, 2, 3 and 4 for each of the test sequences. Meanwhile, Table 5 summarizes the results by averaging the result from table 1, 2, 3, and 4. In each table, the best performance is highlighted in bold. To sum up, in the *CampusRoad* test sequences, we can see that our proposed algorithm outperforms all other methods by having high value in all measurement parameters. In *MovingCurtain* test sequences, Bayesian learning-based is the best in term of accuracy, however, the difference is only by 0.0028 compared with our proposed method. In addition, in this test sequence, our proposed method has the highest precision value and harmonic mean of precision-recall, F_1 scores which make our proposed algorithm surpassed the Bayesian method holistically. Meanwhile, in *WaveBeach* test sequences, uniform-Gaussian[20], performs in most measurement parameters compared with the rest of the studied methods, however, our proposed algorithm still performs the best in term of precision and only have a small difference in term of accuracy. In *Fountain* test sequences, our proposed algorithm scores five best values out of nine test parameters and the highest precision value that is very huge difference compared to other methods.

Based on Table 5, we can summarize that our proposed method is the best in term of accuracy, precision and recall. The consistent high precision value of our proposed algorithm is outstanding and make the algorithm as a reliable detection

Table 1 Quantitative result for MovingCurtain test sequences

Algorithm	Accuracy	Precision	Recall	FNR	FDR	NPV	TNR	F_1 Score	Fallout
Uniform+Gaussian[25]	0.984	0.8131	0.8304	0.1696	0.1869	0.9921	0.8304	0.8217	0.0089
Bayesian[26]	**0.9901**	0.913	0.8016	0.1984	0.087	0.9926	0.8016	0.8537	0.0029
AGMM[27]	0.987	0.9332	0.8019	0.1981	0.0668	0.9895	0.8019	0.8625	0.0031
GMM[28]	0.99	0.9428	0.8237	0.1763	0.0572	0.9919	0.8237	0.8792	**0.0023**
Pfinder[29]	0.9833	0.8007	**0.9605**	**0.0395**	0.1993	**0.9974**	**0.9605**	0.8733	0.0153
Proposed	0.9873	**0.9624**	0.8465	0.1535	**0.0376**	0.9889	0.8465	**0.9007**	0.0024

Table 2 Quantitative Result for WaveBeach test sequences

Algorithm	Accuracy	Precision	Recall	FNR	FDR	NPV	TNR	F_1 Score	Fallout
Uniform+Gaussian[25]	**0.9902**	0.9248	**0.8852**	**0.1148**	0.0752	**0.9937**	**0.8854**	**0.9045**	0.004
Bayesian[26]	0.972	0.8169	0.7378	0.2622	0.1831	0.9817	0.7378	0.7753	0.0116
AGMM[27]	0.986	0.9495	0.7952	0.2048	0.0505	0.9878	0.7952	0.8655	0.0025
GMM[28]	0.9865	0.9519	0.7998	0.2002	0.0483	0.9883	0.7998	0.8692	0.0024
Pfinder[29]	0.9834	0.9498	0.7459	0.2541	0.0502	0.9849	0.7459	0.8356	0.0024
Proposed	0.9888	**0.9636**	0.8067	0.1933	**0.0364**	0.9899	0.8067	0.8782	**1.60E-03**

Table 3 Quantitative Result for CampusRoad test sequences

Algorithm	Accuracy	Precision	Recall	FNR	FDR	NPV	TNR	F_1 Score	Fallout
Uniform+Gaussian[25]	0.9854	0.8204	0.9116	0.0884	0.1796	0.9953	0.9116	0.8636	0.0106
Bayesian[26]	0.9799	0.6868	0.6937	0.3063	0.3132	0.9898	0.6937	0.6902	0.0106
AGMM[27]	0.9408	0.3172	0.285	0.715	0.6828	0.9668	0.285	0.3003	0.0286
GMM[28]	0.9702	0.6927	0.5435	0.4565	0.3073	0.9798	0.5435	0.6091	0.0108
Pfinder[29]	0.9373	0.3688	0.5	0.5	0.6312	0.9753	0.5	0.4245	0.0415
Proposed	**0.9941**	**0.928**	**0.9726**	**0.0274**	**0.072**	**0.9983**	**0.9726**	**0.9498**	**0.0046**

Table 4 Quantitative result for Fountain test sequences

Algorithm	Accuracy	Precision	Recall	FNR	FDR	NPV	TNR	F_1 Score	Fallout
Uniform+Gaussian[25]	0.9858	0.7003	0.5654	0.4346	0.2997	0.9907	0.5654	0.6257	0.0052
Bayesian[26]	0.9814	0.6341	**0.8832**	**0.1168**	0.3659	**0.9964**	**0.8832**	0.7382	0.0156
AGMM[27]	0.9863	0.6686	0.6243	0.3757	0.3314	0.9923	0.6243	0.6457	0.0063
GMM[28]	0.9892	0.7127	0.7167	0.2833	0.2873	0.9945	0.7167	0.7147	0.0056
Pfinder[29]	0.9862	0.7915	0.5209	0.4791	0.2085	0.9891	0.5209	0.6283	0.0032
Proposed	**0.9899**	**0.9041**	0.7954	0.2046	**0.0959**	0.9926	0.7954	**0.8463**	textbf3.10E-03

Table 5 Average of gold standard measurements parameters for all test sequences

Algorithm	Accuracy	Precision	Recall	FNR	FDR	NPV	TNR	F_1 Score	Fallout
Uniform+Gaussian[25]	0.9864	0.8147	0.7982	0.2019	0.1854	**0.993**	0.7982	0.8039	0.0072
Bayesian[26]	0.9809	0.7627	0.7791	0.2209	0.2373	0.9901	0.7791	0.7644	0.0102
AGMM[27]	0.975	0.7171	0.6266	0.3734	0.2829	0.9841	0.6266	0.6685	0.0101
GMM[28]	0.984	0.825	0.7209	0.2791	0.175	0.9886	0.7209	0.7681	0.0053
Pfinder[29]	0.9726	0.7277	0.6818	0.3182	0.2723	0.9867	0.6818	0.6904	0.0156
Proposed	**0.99**	**0.9395**	**0.8553**	**0.1447**	**0.0605**	0.9924	**0.8553**	**0.8938**	**0.0029**

method.

5 Conclusion

This research presented a novel real time based background modeling technique for the scene with varied dynamic background. The proposed algorithm are tested in four complex scenes and compared with five recent algorithm studied. Experimental results show that our proposed algorithm is able to give a consistent, accurate and precise detection results. The common problems in detection method are also successfully handled.

Acknowledgment

This research work is supported by Fundamental Research Grant Scheme (FRGS13-030-0271) under the Ministry of Higher Education of Malaysia (MOHE).

References

[1] M. Piccardi.(2004), "Background subtraction techniques: a review", *in IEEE International Conference on Systems*, Man and Cybernetics, The Hague, Netherlands.

[2] M. Xiao, C. Han and X. Kang. (2006), "A background reconstruction for dynamics scenes", *in IEEE International Conference on Information Fusion*, Florence, Italy.

[3] Z. Hou and C. Han. (2004), "A background reconstruction algorithm based on pixel intensity classification in remote video surveillance system", *in 7th International Conference on Information Fusion*, Stockholm, Sweden.

[4] C. Kamath and S. C. Cheung. (2004), "Robust techniques for background subtraction in urban traffic video", *Proceedings of SPIE*, Vol. 5308, No. 1, pp.881-892.

[5] T. Gao, J. Zhang, W. Gao and Z. Liu. (2009), "A robust technique for background subtraction in traffic video", *in 15th International Conference on Neural Information Processing*, Auckland, New Zealand.

[6] M. H. Hung, J. S. Pan and H. C. Hsieh. (2014), "A fast algorithm of temporal median filter for background subtraction", *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 5, No. 1, pp.33-41.

[7] S. Asif, A. Javed and M. Irfan. (2014), "Human identification on the basis of gaits using time efficient feature extraction and temporal median background subtraction", *International Journal Image, Graphics and Signal Processing*, Vol. 3, No. 2, pp.35-42.

[8] M. Xiao, C. Han and X. Kang. (2006), "A background reconstruction for dynamic scenes", *in IEEEE 9th International Conference on Information Fusion.*

[9] E. L. Rubio and R. M. Baena. (2011), "Stochastic approximation for background modelling", *Computer Vision and Image Understanding*, Vol. 115, No. 6, pp.735-749.

[10] C. Stauffer and W. E. L. Grimson. (1999), "Adaptive background mixture models for real-time tracking", *in IEEE Computer Society Conference*, Colorado, 1999.

[11] S. Mukerjee and K. Das.(2013), "An adaptive gmm approach to background subtraction for application in real time surveillance", *International Journal of Research in Engineering and Technology*, Vol. 2, No. 1, pp.25-29.

[12] M. Nimse, S. Varma and S. Patil. (2014), "Shadow removal using background subtraction and reconstruction", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4, No. 4, pp.324-327.

[13] H. Asaidi, A. Aarab, M. Bellouki. (2014), "Shadow elimination and vehicles classification approaches in traffic video surveillance contex", *Journal of Visual Languages and Computing*, Vol. 12. No.4, pp.333-345.

[14] A. Elgammal, R. Duraiswami, D. Harwood and L. Davis. (2002), "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", *Proceeding of the IEEE*, Vol. 90, No. 7, pp.1151-1162.

[15] J. Lee and M. Park. (2012), "An adaptive background subtraction method based on kernel density estimation", *Journal on the Science and Technology of Sensors and Biosensors*, Vol. 12, No. 9, pp.12279-12300.

[16] J. B. Cruz, A. T. Ali and E. L. Dagless. (1993), "A temporal smoothing technique for real-time motion detection", *in 5th Computer Analysis International Conference*, Budapest, Hungary.

[17] C. Ridder, O. Munkelt and H. Kirchner. (1995), "Adaptive background estimation and foreground detection using kalman filter", *in Proceedings of International Conference on Recent Advances in Mechatronics*, Istanbul, Turkey.

[18] Z. Hou and C. Han. (2004), "A background reconstruction algorithm based on pixel intensity classification in remote video surveillance system", *Proceedings of the Seventh International Conference on Information Fusion*, Vol. 2.

[19] L. Cao and Y. Jiang. (2013), "An effective background reconstruction method for video objects detection", *in International Conference on Networking and Distributed Computing.*

[20] N. Chaki , S. H. Shaikh and K. Saeed. (2014), *Exploring Image Binarization Techniques*, Springer.

[21] T. Fawcett. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27, pp.861-874.

[22] D. Powers. (2007), *Evaluation: from precision, recall and f-factor to roc, informedness, markedness* & *correlation*, Flinders University of South Australia, Adelaide.

[23] K. Kumar and S. Agarwal. (2013), "An efficient hierarchical approach for background subtraction and shadow removal using adaptive GMM and color discrimination", *International Journal of Computer Applications*, Vol. 75, No.7.

[24] L. Li, W. Huang, I. Y.-H. Gu and Q. Tian. (2004), "Statistical modeling of complex backgrounds for foreground object detection",*IEEE Transactions on Image Processing*, Vol. 13, No. 11, pp.1459-1472.

[25] Z. Zivkovic and F. V. d. Heijden. (2006), "Efficient adaptive density estimation per image pixel for the task of background subtraction",*Pattern Recognition Letters*, Vol. 27, No. 7,

[26] C. Stauffer and W. Grimson.(2000), "Learning patterns of activity using real-time tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp.747-757.

[27] C. Wren, A. Azarbayejani, T. Darrell and A. Pentl. (1997), "Pfinder: real-time tracking of the human body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.780-785.

Corresponding author
N.A. Zainuddin can be contacted at: fiqahzainuddin@gmail.com