EasyChair Preprint

№ 5335

Performance Evaluation of Background Subtraction Techniques for Video Frames

Salman Qasim, Kaleem Nawaz Khan, Miao Yu and Muhammad Salman Khan

April 18, 2021

Performance Evaluation of Background Subtraction Techniques for Video Frames

Salman Qasim
Department of Computer Science & IT
University of Engineering and Technology
Peshawar, Pakistan
salmanqasim@uetpeshawar.edu.pk

Kaleem Nawaz Khan
Artificial Intelligence in Healthcare,
IIPL, National Center of Artificial Intellgience
University of Engineering and Technology
Peshawar, Pakistan
kaleemnawaz@uetmardan.edu.pk

Miao Yu
School of Computer Science
University of Lincoln
myu@lincoln.ac.uk

Muhammad Salman Khan
Department of Electrical Engineering
Artificial Intelligence in Healthcare,
IIPL, National Center of Artificial Intellgience
Univerity of Engineering and Technology
Jalozai campus, Peshawar, Pakistan
salmankhan@uetpeshawar.edu.pk

***Abstract*—The fundamental working of background subtraction is to identify the moving region by taking pixel-wise difference of the current frame from the previous one. The proposed study presents the comparison and implementation of different background subtraction techniques i.e., frame-difference method, mixture of Gaussian model 2 (MOG2) and k-nearest neighbor (KNN) for background subtraction. For all the three techniques, prior to segmentation, background modeling and then features extraction steps are performed. It is investigated that on the same dataset, frame-difference technique outperforms both MOG2 and KNN and achieve accuracy of 89.98%, recall of 94.43% precision 79.55% and F1-score of 81.42%.**

***Index Terms*—background subtraction, frame-difference, mixture of Gaussian model 2, k-nearest neighbor, features extraction, segmentation**

I. Introduction

In computer vision, to understand the analysis of video sequence it is important to know the underlying concepts and working principles. In different applications like surveillance video, object and vehicle detection, the main step is identifying the object which is moving. Thus, the important operation is to separate the moving objects which is known as "Foreground" from the fixed information which is known as "Background" [1]. The automatic recognition system plays a significant role in the detection of moving objects, extraction of the features, human body detection and many other applications. The number of vehicles rapidly increasing in non-rural and national road system rose the requirement for successful checking and the management of street traffic. In the upcoming era, it is anticipated that about 3.7 million miles of streets and roads are evaluated to increment by 30%. Particularly the detection of huge number of vehicles which consists of not only small but large vehicles as well, is not an easy task [2]. Similarly for elderly care falling is a dangerous activity, which seriously affects an elderly person's health and commonly happen among the old people community. According to the Public Health England, those individuals who are 65 and above are likely to fall once a year is assessed to be 30% and for those elderly people who are 80 or above that, it is estimated that it exceeds to 50% [3].

For detection and finding of the moving objects in an automatic surveillance, object tracking and traffic examination, background subtraction algorithm is the most commonly used approach. Similarly due to some limitations in detection of fall methods that has been presented for elderly care, the fall detection systems based on CV, becomes the mainstream one, which uses the ordinary RGB camera and background subtraction techniques are used to extract the human silhouette features from an input video stream that are extracted by convolutional neural network to detect a fall. In this literature, different methods have been reviewed and developed for video analysis. These strategies point on evaluating the picture arrangements properly. Essentially, to recognize and find the moving objects, and then to track them. Furthermore, several methods have also been presented for the designing of the background. A correlation study in between these techniques is acknowledged so as to enhance the features and coherence of every technique for extraction of pixels and pursue the specific situation of moving objects. A comparative evaluation of three different techniques for the extraction of foreground is presented in this article to distinguish between moving and non-moving objects, the techniques are: frame difference, mixture of Gaussian and KNN background subtraction.

The paper is organized in six different sections. After the introduction to the problem in section I, section II presents the detailed literature review and existing methods. Proposed methods are discussed in section III. Moreover, section IV

presents the experiments and results obtained. The results comparison and analysis are discussed in section V. Finally the last section conclude the paper and provides future directions.

II. Literature Review

For the detection of the foreground objects, and also the modelling of the background, background subtraction is an important method. Many studies can be found on how the subtraction of the background plays an important role in computer vision. Different techniques have been presented to deal with background subtraction issue. E. Canayaz and V. G. Bocekci, proposed by using surveillance videos, heavy vehicles can be detected using the Gaussian Mixture Model, Approximate Median Filtering and Inter Frame Difference for subtraction of background and to compare the performance [2]. Furthermore, after the removal of background image from the videos, blob analysis and morphological opening was applied and the blob having minimum area of detected object within the frame, detection of the heavy vehicles was achieved. Various background subtraction techniques gave different results, and those results were examined. The results were dependable with comparison of performance which showed that Gaussian mixture model was durable in real time air tracker in any changing outside environment. Although in this article, different techniques gave simultaneous results according to the performance and had distinctive superiority among them still Gaussian Mixture Model detected way more efficiently.

For the elder people community, Yu Miao et al. proposed computer vision based fall detection system [3]. Background subtraction is applied from a recorded video stream which is used to extract the human silhouette region. The silhouettes which are extracted are pre-processed and then exercised to train a CNN, that is further utilized for classification of postures and detection of a fall. Results showed that suggested model gives excellent performance than the classic ones as well as achieves a tremendous fall detection accuracy. The only miscalculation was of the bend posture which was also detected as a fall. Similarly, S. Mohamed, et al. proposed Mixture of Gaussian (MOG) method. MOG has low rate of suitability, complexity and memory consumption and to detect the object for the outdoor environment. This algorithm is way more adaptive and robust in background subtraction method and can handle multi-modal distributions [4].

In [5], the authors specifically evaluates the correlative assessment of Object Tracking with subtraction of background techniques which are running average Gaussian (RAG), eigen background (EB) and frame differencing (FD). Statistical analysis, confusion matrix, is utilized to assess every one of the techniques. By using CDnet datasets, evaluation was performed on realistic outdoor environment. FD and EB showed similar results whereas for RAG the performance was poor.

Benraya and N. Benblidia, proposed the most known background subtraction techniques which are generally utilized for different vision related tasks [6]. This is accomplished by methods for the mask abstraction nature of the forefront utilizing these techniques. Comparative survey between these techniques is acknowledged so as to improve the nature and ability of every technique in extricating pixels and following the specific situation of object. A correlation investigation of four techniques of foreground abstraction is received in this study to separate the moving object from the stationary ones the techniques are: MOG2, MOG, KNN and Geometric Multigrid (GMG). The preliminary results showed 8 assessment measures to look at the extraction nature of the techniques, results obtained from GMG were best one, the results of MOG were also satisfying, results gathered from MOG2 and KNN were not satisfying.

Different deep learning based methods are also utilized in the literature [7] for the implementation of BS algorithms, but due to some of their limitations [8], the main focus is on classical BS algorithms. They require a large amount of dataset to perform better than other techniques. For the training of model in a deep learning environment, expensive GPUs and machines are used which increases the cost for the user. Based on just mere learning and classifiers, it is not easy to understand the output results. For such tasks Convolutional Neural Networks are used. On contrary, the main advantage of the algorithms used for the background subtraction is that they are robust against the movement of the background, for instance the moving branches of a tree and leaves. Without destroying the existing background model, objects are allowed to become a part of the background.

Frame difference BS has different benefits which includes: objects with uniformly distributed intensity, its computationally cheap and highly adaptive background model. Similarly, the MOG2 method tracks multiple Gaussian distributions simultaneously. MOG2 maintains a density function for each pixel. It is capable of handling multi-modal background distributions. Since MOG is parametric, the model parameters can be adaptively updated without keeping a large buffer of video frames. With the implementation of KNN BS it is simple and easy to use, also easy to understand and can be used for classification or regression. Not sensitive to outliers.

In [9], the authors discussed the improvement and use of moving objects location based on rapid video. An enhanced detection of object method dependent on subtraction of background and frame difference is presented. Also, morphological filter and threshold area de-noising is engaged for the post processing of image for moving object. Obtained results showed that the presented study detected the movement of object accurately and adequately. The proposed methodology and relevant steps are discussed in the following section.

III. Methodology

In this section a comparison of various background subtraction techniques will be performed that can be used in different domains such as object tracking and fall detection [10]. The aforementioned techniques are:

- Frame Difference Background Subtraction
- Mixture of Gaussian (MOG2)
- KNN Background Subtraction

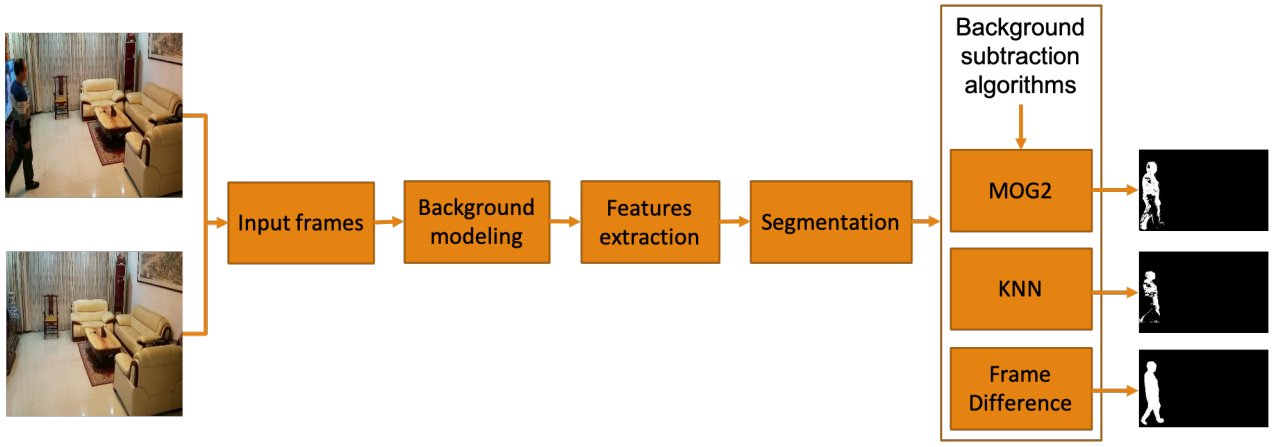

Fig. 1: Proposed Methodology

Fig. 2: Frame Difference Background Subtraction

The proposed methodology is depicted in Fig. 1. Almost all the background subtraction algorithms are made up of different processing components that are described in following subsections:

A. Background Modelling

For the background subtraction algorithm, the background model is necessary. For the incoming video frames, background modelling can be used as a reference. Moreover, the background model has a significant job since video frames are ordinarily not totally liberated from foreground during the starting stage [11]. As an outcome, the model gets ruined by foreground objects away from background model, which gives incorrect classifications.

$$D_k(x,y) = \begin{cases} 1 \\ \quad if|f_k(x,y) - f_{k-1}(x,y)| > T \\ 0 \end{cases} \tag{1}$$

B. Feature Extraction

The relevant information which is represented by the adequate features, must be selected, so to compare the video frames with the background image. RGB and gray scale intensities are used as features by many algorithms. In few of the cases, intensity of pixels and some other features are joined. Moreover, it is important to select choice of feature region. Features can be extracted over the blocks, patterns or pixels. Features that are pixel-wise mostly yield segmentation result's that are noisy since they don't encode local relation, while the pattern-wise and the block-wise features likely to be indifferent to minimum changes.

C. Segmentation

Video frames can be processed with the help of a background model. By extracting the features from corresponding pixels, background segmentation can be performed or region of the pixels of both the frames and utilizing an extension range, such as the Euclidean distance, to calculate the similarities between the pixels. With the similarity threshold and after being compared, each pixel is either labeled as foreground or background. The formation of the overall background subtraction system is formed by the combination of those building blocks. The strength of the system is constantly relying and bounded by the performance of every individual block, i.e., it can not be expected to perform well if one module delivers poor performance. Background subtraction is a very vast field, in this way there exists a number of algorithms for this purpose.

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t}.\eta(X_t, \mu_{i,t} \sum i.t) \tag{2}$$

D. Frame Difference

This technique is achieved by taking the difference between two pictures to decide the existence of moving objects. It can be said that it is the most easiest form of background subtraction. Frame difference, otherwise called temporal difference, utilizes the video outline at time t-1 as the background model for frame at time t [12]. For a given video sample, take the first frame and then find the absolute difference with another frame. First, by taking the difference of the corresponding pixels of the k frame and the k-1 frame using Equation 1, a binary difference image is obtained. T represents the threshold. In the binary image, foreground points are considered as one-value pixels while background points are considered as zero-value pixels [13]. Fig. 2 shows the result of frame difference background subtraction.

E. MOG2

One of the most common and popular background subtraction technique is the Gaussian Mixture Model. Gaussian

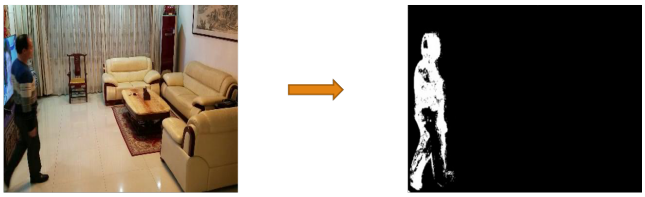

Fig. 3: MOG2 Background Subtraction

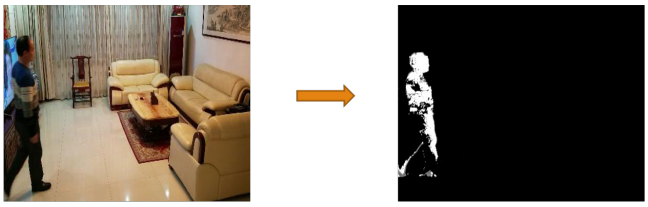

Fig. 4: KNN Background Subtraction

Mixture based algorithm is used for segmentation of foreground/background. Stauffer and Grimson [14] proposed a scheme for the representation of background which is pixel-wise by the usage of Mixture of Gaussian (MOG) and the updating background to upgrade the variance and intensity of the mean for each pixel in real-time. Based on the learning factor, the parameters of the model are updated. The least probable model is eliminated If no match is found and with the current pixel values replaced by a new Gaussian [15]. For moving background scene having multiple background variations, the MOG based methods are effective but they are sensitive to illumination changes and noise [16].

In [17], by mixing Gaussian the authors modeled the background, so to find a match at a particular location, where every pixel is compared to existing models. Given by the formula in Equation 2, the probability of observing the current pixel value is measured in a multidimensional case. Where K is the

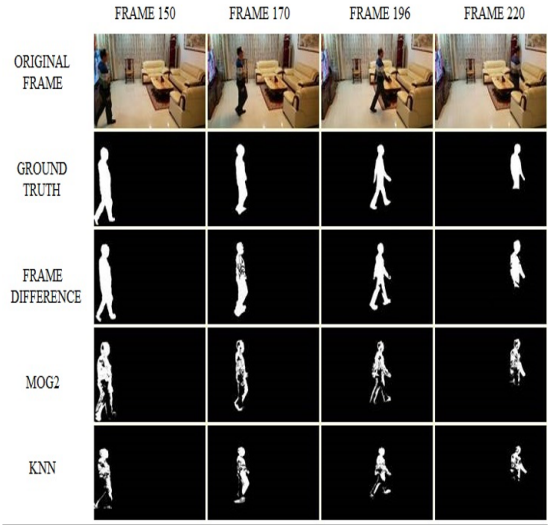

Fig. 5: Comparison of different background subtraction techniques

TABLE I: Evaluation metrics for video frame 150

Methods	*Accuracy*	*Recall*	*Precision*	*F1-Score*
Frame Difference	0.9117	0.8878	0.8678	0.8772
MOG2	0.7144	0.6299	0.7108	0.6679
KNN	0.4994	0.3804	0.8991	0.5346

TABLE II: Evaluation metrics for video frame 170

Methods	*Accuracy*	*Recall*	*Precision*	*F1-Score*
Frame Difference	0.9044	0.9660	0.8902	0.9266
MOG2	0.7077	0.5610	0.8254	0.6680
KNN	0.7001	0.4523	0.8327	0.5839

no. of distributions, $\omega_{i,t}$ weight linked to the ith Gaussian at time t having mean $\mu_{i,t}$ and $\sum i.t.\ \eta$ standard deviation, is probability density function.

MOG2, a new and updated version of MOG applies the same approach as in older MOG with some added features. For each pixel, the convenient number of Gaussian distribution is selected by its own. Option of selection of shadow to be detected or not is also present. Due to changes in illumination, it gives better adaptability to different scenes [18].

Sometimes, in a background scenery there are often some non static objects like branches and leaves of trees, which are showing movements due to the wind. This type of background movement shows the pixel intensity varies considerably, therefore in this type of situation, the representation of pixel intensity will not be considered by a simple Gaussian. Fig. 3 represents the results of MOG2 background subtraction technique.

F. KNN Background Subtraction

As the name indicates, it is a KNN based background/ foreground segmentation algorithm. Loftsfarden and Quesenberry in 1965 presented density estimation method [19], which is also known as the KNN method and is more efficient for local density estimation. The density estimation formula approximately is given in Equation 3.

$$[t]p(x|x_i) \approx \frac{1}{NV} \sum_{m=1}^{N} b^m K \left(\frac{||x_i - x||}{D} \right) \tag{3}$$

Here, K is the kernel function, subject to uniform distribution. If $u < 1/2$, then the kernel $K(u) = 1$, otherwise 0. If the video sequence sample is assigned to foreground, the value of b^m is 0. The background model only deals with samples that satisfy b^m and are classified as background. As it can be seen in Equation 4, where, if $P(x|x_i)$ is greater than a certain threshold of T, the pixel is considered the background. And the choice of T is closely related to the value of V. Fig. 4 shows output result of the KNN background subtraction method.

$$[t]M_i = \begin{cases} \text{Foreground, } if P(x|x_i) < T \\ \text{Background, } otherwise \end{cases} \tag{4}$$

TABLE III: Evaluation metrics for video frame 196

Methods	*Accuracy*	*Recall*	*Precision*	*F1-Score*
Frame Difference	0.8801	0.9614	0.8787	0.9182
MOG2	0.8555	0.5490	0.7999	0.6509
KNN	0.8500	0.3967	0.8074	0.5321

TABLE IV: Evaluation metrics for video frame 222

Methods	*Accuracy*	*Recall*	*Precision*	*F1-Score*
Frame Difference	0.8998	0.9443	0.7955	0.8142
MOG2	0.8380	0.5987	0.8354	0.7177
KNN	0.5001	0.4050	0.7996	0.5376

IV. Experimental Setup

This section focuses on the comparison of background subtraction techniques using methods (Frame difference, MOG2 and KNN) in video captured with a single person as a moving object. A short video of about 2 minutes was played and a total of 110 different frames of a video file were captured. The target is to compare the extracted features with each technique used, and the image obtained after the processing of the algorithm is then compared with the given ground truths. There are different frames of a video file containing a single person which acts a moving object and manually annotation of those input frames are the ground truth of those images and will be compared with the background subtraction techniques. Out of these frames 4 random frames are selected that are shown in Fig. 5.

The analysis was carried out on Intel Core m3 CPU with 1.61 GHz processor having 8GB RAM. The video file was recorded with a camera of frame resolution (480 x 270) pixel in mp4 format. Python OpenCV library was used for the experimentation on video file.

V. Results

The evaluation of the video constitutes basically consist of some factors which are true positives (TP), false positives (FP), false negatives (FN) and false positive (FP). From these values we calculate Accuracy, Precision, Recall and F1 score using the Equations (5) to (8). All the pixels are therefore classified according to four categories:

- TP (true positives): foreground was detected as foreground
- FP (false positives): background was detected as foreground
- TN (true negatives): background was detected as background
- FN (false negatives): foreground was detected as background

$$\text{Sensitivity/Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Specificity/Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

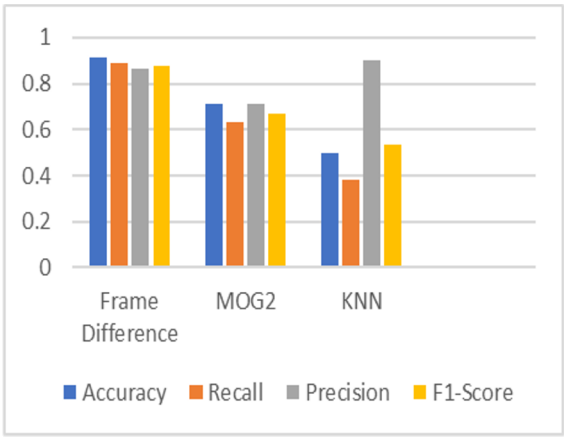

Fig. 6: Evaluation metrics Comparison for video frame 150

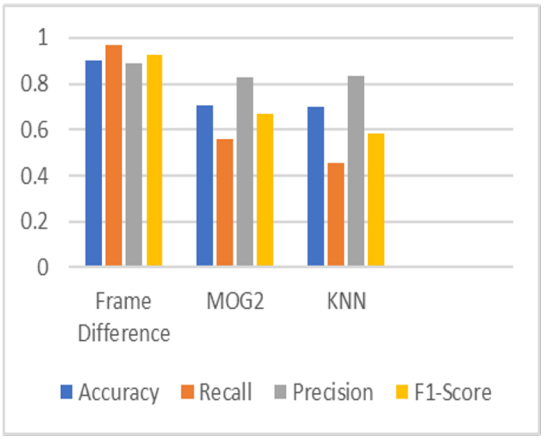

Fig. 7: Evaluation metrics Comparison for video frame 170

$$\text{F-score} = 2\frac{Recall \times Precision}{Recall + Precision} \tag{8}$$

If the foreground pixel was detected as a foreground than the foreground pixel detected as background, than a high TP will be obtained (TP $>$ FN). If the value of FP is small, it shows that the background pixels were detected as background. For best results, it is significant to have high TN and FP should be low. The technique having high TP and low FP will be considered best for background subtraction. Table (1) to (4) and Figure (6) to (9) shows performance comparison of all the three background subtraction techniques for 4 different video frames.

VI. Conclusions

This study presented the comparison and implementation of different background subtraction techniques i.e., frame-difference, MOG2 and KNN background subtraction that are extensively being used in field of computer vision. It is accomplished by methods for extraction of silhouette regions of foreground from given video sample. We performed some experiments and got the performance evaluation of the aforementioned techniques. Frame difference method and MOG2

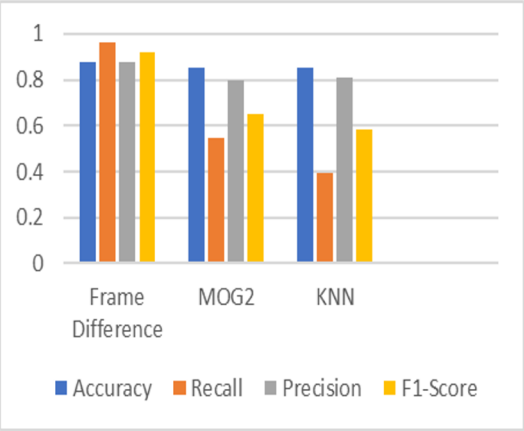

Fig. 8: Evaluation metrics Comparison for video frame 196

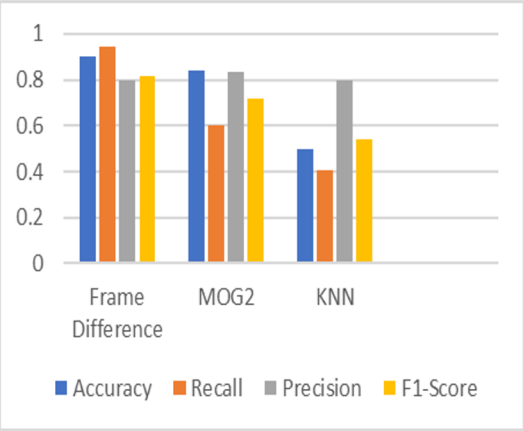

Fig. 9: Evaluation metrics Comparison for video frame 222

shows good results on background subtraction tasks and achieved accuracies of 89.98% and 83.80%, respectively. KNN didn't perform well as compared to the other two methods and achieved 50% score. For future work, this study can also be implemented with some machine learning or deep learning algorithms for obtaining better accuracy for any large datasets.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of Artificial Intelligence in Healthcare, Intelligent Information Processing Lab, National Center of Artificial Intelligence, UET Peshawar for the necessary research infrastructure and support we are also grateful of NVIDIA Corporation for supporting this research by providing Titan X Pascal high computing facility

REFERENCES

[1] Neha S Sakpal and Manoj Sabnis. Adaptive background subtraction in images. In *2018 International Conference on Advances in Communication and Computing Technology (ICACCT)*, pages 439–444. IEEE, 2018.

[2] Emre Canayaz and Veysel Gökhan Böcekçı. Comparison of performance of different background subtraction methods for detection of heavy vehicles. In *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 50–54. IEEE, 2018.

[3] Miao Yu, Liyun Gong, and Stefanos Kollias. Computer vision based fall detection by a convolutional neural network. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 416–420, 2017.

[4] Shahrizat Shaik Mohamed, Nooritawati Md Tahir, and Ramli Adnan. Background modelling and background subtraction performance for object detection. In *2010 6th International Colloquium on Signal Processing & its Applications*, pages 1–6. IEEE, 2010.

[5] Dennis Aprilla Christie and Topan Sukma. Comparative evaluation of object tracking with background subtraction methods. In *2018 Third International Conference on Informatics and Computing (ICIC)*, pages 1–6. IEEE, 2018.

[6] Imane Benraya and Nadjia Benblidia. Comparison of background subtraction methods. In *2018 International Conference on Applied Smart Systems (ICASS)*, pages 1–5. IEEE, 2018.

[7] Chulyeon Kim, Jiyoung Lee, Taekjin Han, and Young-Min Kim. A hybrid framework combining background subtraction and deep neural networks for rapid person detection. *Journal of Big Data*, 5(1):22, 2018.

[8] Thierry Bouwmans, Sajid Javed, Maryam Sultana, and Soon Ki Jung. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *Neural Networks*, 117:8–66, 2019.

[9] Wei Wan, Shoujun Tang, and Hongyang Zhang. Moving object detection based on high-speed video sequence images. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, pages 906–910. IEEE, 2019.

[10] Muhammad Salman Khan, Miao Yu, Pengming Feng, Liang Wang, and Jonathon Chambers. An unsupervised acoustic fall detection system using source separation for sound interference suppression. *Signal processing*, 110:199–210, 2015.

[11] Mohammadreza Babaee, Duc Tung Dinh, and Gerhard Rigoll. A deep convolutional neural network for video sequence background subtraction. *Pattern Recognition*, 76:635–649, 2018.

[12] Pritee Gupta, Yashpal Singh, and Manoj Gupt. Moving object detection using frame difference, background subtraction and sobs for video surveillance application. In *The Proceedings of the 3rd International Conference System Modeling and Advancement in Research Trends*, 2014.

[13] Honghai Liu and Xianghua Hou. Moving detection research of background frame difference based on gaussian model. In *2012 International Conference on Computer Science and Service System*, pages 258–261. IEEE, 2012.

[14] Chris Stauffer and W Eric 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)*, volume 2, pages 246–252. IEEE, 1999.

[15] Abdourahman Houssein Ahmed, Kidiyo Kpalma, and Abdoulkader Osman Guedi. Human detection using hog-svm, mixture of gaussian and background contours subtraction. In *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 334–338. IEEE, 2017.

[16] Han-Hui Hsiao and Jin-Jang Leou. Background initialization and foreground segmentation for bootstrapping video sequences. *EURASIP Journal on Image and Video Processing*, 2013(1):12, 2013.

[17] Chris Stauffer and W. Eric L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):747–757, 2000.

[18] Zoran Zivkovic and Ferdinand Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780, 2006.

[19] Parisa Darvish Zadeh Varcheie, Michael Sills-Lavoie, and Guillaume-Alexandre Bilodeau. A multiscale region-based motion detection and background subtraction algorithm. *Sensors*, 10(2):1041–1061, 2010.