



ABGS Segmenter: pixel wise adaptive background subtraction and intensity ratio based shadow removal approach for moving object detection

Supriya Agrawal¹ · Prachi Natu¹

Accepted: 21 November 2022 / Published online: 9 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Background subtraction approaches are used to detect moving objects with a high recognition rate and less computation time. These methods face two challenges: selecting the appropriate threshold value and removing shadow pixels for correct foreground detection. In this paper, we solve these challenges by proposing a new background subtraction method called ABGS Segmenter, which is based on a two-level adaptive thresholding approach where a reference frame is created using mean-based thresholding to generate the initial value of the threshold and accelerates the process of foreground segmentation for remaining frames by adaptively updating the threshold value at the pixel level. ABGS Segmenter is also capable of removing shadow pixels by fusing the chromaticity-based YCbCr color space model with the intensity ratio method for improving the percentage of correct pixels' classification measure. Comprehensive experiments are evaluated on three benchmark datasets (Highway, PETS 2006, and SBU) and observed that the proposed work achieves better results than existing methods.

Keyword Adaptive threshold · Background subtraction · Color space model · Shadow detection · Shadow removal

1 Introduction

Traditional video surveillance systems are used to monitor activities like traffic control, people detection, crime detection, and many more. These systems are dependent on human operators. It has been observed that human-based surveillance systems have certain limitations, such as the inability to monitor multiple screens at the same time or track an object over a long period [1]. To overcome these issues,

✉ Supriya Agrawal
supriya.agrawal@nmims.edu

¹ Department of Computer Engineering, NMIMS University, Mumbai, India

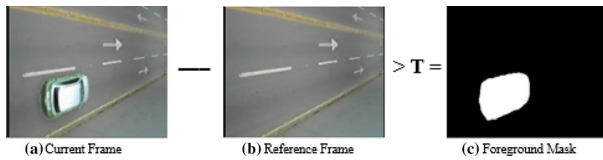
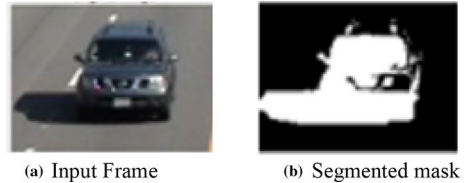


Fig. 1 Basic approach of the Background Subtraction method

Fig. 2 **a** Input frame contains an object with shadow **b** Segmented mask of foreground extraction with shadow



intelligent surveillance systems have been popular in recent years by providing a high level of security to analyze the captured videos efficiently. To analyze video frames, the object segmentation approach is employed to precisely identify the moving object for recognition. The most widely used video segmentation approach is known as Background Subtraction [2, 3]. It is used to detect moving objects in video sequences by calculating the pixel-wise difference between the current frame and reference frame and then comparing the difference with the threshold value (T), to generate a binary mask, as shown in Fig. 1 [3]. With the help of the threshold value, pixels are divided into two classes: foreground pixels are considered moving parts of the frame, and background pixels are considered stationary parts of the frame [4].

Background Subtraction (BS) techniques become more challenging when they require choosing a threshold value for foreground segmentation. Many methods have been developed to detect moving objects using background subtraction approaches from video sequences, but segmented results are not great [5–7]. Basic background subtraction models rely on either manual selection of threshold value or global threshold value, producing unsatisfactory results in terms of correct pixels' classification while generating the binary mask. Unfortunately, these methods don't always lead to good results, especially when moving objects stop moving or when the background is changing [7, 8].

To solve this issue, statistical models were used for background subtraction by removing the dependency on a manual threshold value [9–11]. These statistical models were integrated with multi-model distributions to calculate the learning rate parameter for threshold selection. However, these models are reliable when there are sudden illumination changes in dynamic scenes, but threshold selection is dependent on tuning the learning rate parameter, which is time-consuming. Also, these background subtraction methods produce unsatisfactory results in the extraction of correct foreground objects due to the existence of shadows (as shown in Fig. 2 [12]), which cause the distorted geometrical shape of the moving objects.

To solve the problem of manual or global thresholding, a radius-based threshold value method was proposed to detect foreground objects[13], but a few pixels were

misclassified due to shadow. Shadows follow the same movement patterns as foreground objects and exhibit a comparable amount of intensity change, which leads them to become a part of the foreground. There are numerous shadow detection and removal techniques available in the literature to attain good accuracy [14, 15], but they have limitations, including (a) detecting shadows in moving objects using color feature segmentation, but it is unable to discriminate them from black objects [16], and (b) shadows are based on a darker intensity region, but it produces a deformed shape of the object [17]. Moreover, these techniques produce less accuracy by generating many false positives.

Existing methods achieve good results. However, it becomes challenging to address the problems of threshold value selection and shadow pixel removal for accurate foreground segmentation. In this research, we try to solve the mentioned issues from a new perspective and propose a framework called ABGS Segmenter: Adaptive Background Subtraction and Shadow removal. The main aim of the proposed ABGS segmenter is to predict pixel-wise adaptive threshold values in each frame for foreground and background segmentation and remove the shadow pixel (if it exists) to improve the accuracy of object segmentation. Traditional background subtraction methods are sensitive to shadow appearances in frames, which affect the performance of the developed background subtraction model by misclassifying shadow pixels as foreground pixels. Fortunately, chromaticity-based shadow removal shows a great impact on this factor. Therefore, we have combined adaptive background subtraction and chromaticity-based shadow removal approaches to generate a reliable method for accurate foreground detection in static scenes using a single steady camera.

Our proposed framework consists of two stages: (i) background subtraction and (ii) shadow removal, which are inspired by the algorithms proposed in [18, 19]. In [18], the pixel-wise background subtraction method is proposed where parameters are adjusted at run time and the background is updated using a learning rate. This method has multiple parameters it takes time to tune the parameters. On the other hand, the algorithm proposed in [19] is used for shadow removal, where the local average color space model is used to remove the shadow pixels. However, this technique cannot remove outdoor shadow pixels accurately, which leads to deform the shape of the objects and generate many false positives. Background subtraction and shadow removal methods have performed well individually, but few background subtraction methods do not perform well if segmented foreground results are associated with shadows.

With this motivation, our proposed framework integrates an adaptive background subtraction method with a shadow removal model to detect foreground moving objects correctly. Here, we have proposed two-level adaptive thresholding in which the initial threshold is calculated using the mean value of the first frame to create the reference frame, and then the threshold value is updated automatically using a statistical approach for other frames at the pixel level for foreground extraction. The reference frame is updated to check the appearance of the objects when they appear first time in the frame and consider the objects as part of the background if they do not move for a longer time. This mechanism helps in calculating adaptive threshold values at run time. To obtain shadow-free frames,

our proposed framework is also able to detect the shadow region and remove the shadow pixels using the YCbCr color space model and proposed intensity ratio model. Our method is also capable to remove single and multiple shadows in the frame.

The contributions of the proposed work are as follows:

1. We propose a pixel-wise two-level adaptive thresholding approach to build a background subtraction model to generate a binary mask for foreground detection, which solves the problem of choosing a manual threshold value. The update mechanism of the reference frame of the proposed work aids in recognizing objects that remain stationary for an extended time.
2. We propose a chromaticity-based YCbCr color space model with an intensity ratio method to know the statistics of pixels' intensity for shadow region detection and subsequently remove shadow pixels to reduce false positive rates.
3. We enhance our framework by incorporating morphological operations and connected components to fill the interior gaps in the foreground mask for correct pixel classification.
4. We conduct comparative experiments and evaluate the performance of the proposed framework on benchmark datasets PETS 2006, Highway, and SBU with state-of-the-art methods.

The organization of the paper is as follows: Section 2 discusses related work of background subtraction and shadow removal. Section 3 describes the proposed methodology. Section 4 demonstrates experimental results and comparative analysis. Finally, Section 5 presents the conclusion of the proposed method.

2 Related work

Moving object detection is a key step of video surveillance applications in the computer vision field. In past years, numerous background subtraction approaches have been proposed to detect foreground objects from the video. Sometimes, detected foreground objects are captured with shadows, affecting the performance of intelligent surveillance systems. So, it is highly important to remove shadows from extracted foreground objects. This section discusses various existing methods related to (i) background subtraction models and (ii) shadow detection and removal methods for foreground extraction.

Background Subtraction Methods: To detect foreground objects in video sequences, many background subtraction methods have been presented in the last few decades. The following stages are included in the process of background subtraction process: (1) initializing the background model from N training frames to generate the first background. (2) foreground detection is accomplished by comparing the background image to the current image and classifying pixels as foreground or background (3) background maintenance is used to generate a

foreground mask by updating the background model. Background subtraction methods are not limited and many review papers have been published by providing detailed information on various background subtraction approaches and their applications [2, 6, 15].

Traditional pixel-based background subtraction methods distribute each pixel in a frame locally. In these models, the pixel-wise difference is calculated between the background frame and the current frame using a manual choosing threshold value to classify pixels belonging to foreground or background [20]. In [21], the background subtraction method is proposed using the median filter to detect pedestrians based on their walking manner by pixel-by-pixel comparing the reference and current images, and then applying a median filter to the resulting image to remove noise. When pedestrians wear the same color of the dress as the background images, this method fails, and few pedestrians are detected as non-pedestrian. An adaptive background subtraction model using histogram-based thresholding is proposed in [22] to detect foreground objects in a dynamic scene. This method fails if the moving object is slow which leads to the wrong counting of objects. Hassan et.al. [23] proposed a self-adaptive background subtraction model using histogram analysis to detect a person in a crowded scene. However, these approaches are easy to implement but not able to generate good results if the threshold value is not selected properly.

To overcome the problems of traditional methods, advanced pixels-based statistical methods were proposed which are suitable to generate foreground masks in static and dynamic scenes. Gaussian Mixture method (GMM) is one of the most famous statistical approaches to detecting moving objects in static and dynamic environments. A statistical model using single Gaussian distribution is introduced to recognize people in video sequences [24]. Due to the slow updating rate of the backdrop model, a single Gaussian model could not quickly deal with a dynamic background. To overcome this issue, an adaptive Gaussian mixture-based method is proposed where every pixel is introduced with a mixture of K Gaussian functions and it works well in the dynamic environment [9]. To make it more advanced, adaptive GMM with the Expectation–Maximization method is introduced for background subtraction which is used to predict per-pixel background distributions [25]. A hybrid GMM model was proposed in [26] for a static and dynamic background but not able to handle shadows in the image when the illumination changes. Another statistical background subtraction model is proposed [27] which is integrated with wavelet transform to detect moving objects in illumination changes but misclassify a few pixels when shadow exists. Other mixture models fail in the fast-changing background with Gaussians values ranging between 3 and 5, which causes issues in complex scene detection [28].

There are a few methods proposed to solve issues of fast-changing background environment and the occurrence of shadows [29, 30]. These models estimate the background based on pixel probability distribution in each frame. The clustering model is also introduced called codebook [31] which obtains a multimodel background model in which a codebook for every pixel has been generated to store code-word based on training frames. The codebook method is learned with long video training frames over a limited amount of memory. Further improvements are done by incorporating the spatial and temporal context of each pixel in every frame [32].

Sometimes, codebook methods fail while updating the background model if structural changes occur in moving objects due to noise in video sequences. A well-known visual background extractor (ViBe) method [33] is proposed which handles noisy videos. The key innovation of ViBe is selecting some pixels for background initialization and updating the model from the second frame. This method produces good results in a dynamic environment but failed in a dark background, shadows, and sudden illumination changes in the captured scene. Another method called Pixel Based Adaptive Segmenter (PBAS) is developed which creates a background model taking a sample of observed pixels and the foreground was detected using pixel-based threshold value [18]. To extend this work, a method is proposed in [34] works on weighted directional background subtraction model to improve the segmentation task in static and dynamic scene.

Hybrid models have been discussed in literature which achieve good results in a dynamic environment, shadow removal, and illumination changes. A hybrid model using Kernel Density Estimation and GMM are proposed by [28] which discusses foreground identification and removal of shadow pixels using probability density functions (PDFs) for moving and stationary objects. Subsequently, a Codebook method is proposed [35] in conjunction with the Gaussian Mixture Model (GMM) to detect outliers using mean vectors and covariance matrices. This procedure is also effective in removing shadows. In [36], the wavelet transform method is combined with adaptive GMM to create the background model but it is not robust under sudden illumination changes. The methods proposed in [36, 37] propose a hybrid background subtraction with mean pixels to generate foreground pixels in a real-time environment. The summary of various background subtraction methods and the challenges solved by them is given in Table 1.

Shadow Detection and Removal Methods: Video surveillance systems need to be robust if any changes (like the occurrence of shadows) in the foreground scene cause false detection of the objects. Shadow pixels are occasionally misclassified as object portions, resulting in errors in object localization, segmentation, and tracking[38]. Although detecting shadow regions with the human eye is relatively simple, it is a difficult challenge for a computer to solve when shadow pixels are considered as part of moving regions. To solve such issues, the current research focuses on the detection and removal of shadows. Shadow can be classified as (i) cast shadow and (ii) self-shadow. Shadow is cast when one object is blocked by the light source of another object. Self-shadow is generated when light is not directly reflected over the object. In the literature, many methods are available for the detection and removal of shadows as discussed. The method proposed in [39] discusses two features of shadow detection: a) intensity and b) chromaticity. The choice of features has a great impact on shadow detection results. Shadow regions are different from background regions. Intensity feature-based shadow detection methods predict the intensity range of shadow regions after estimating the background subtraction. These methods sometimes fail when the object and region of shadow have the same color. Earlier, shadow detection systems rely on manual hand-crafting relevant features to distinguish the shadow from the background scene. In the paper [40], a graph-based technique is adopted to detect shadows using intensity features. A Markov Random Field (MRF) method is proposed

Table 1 Challenges addressed using background subtraction approaches

Challenges addressed	Approaches	Captured scenes	Detected objects	Challenges solved/partially solved
Illumination changes	Median filter [21]	Outdoor	Pedestrian	Partially solved
	Histogram model [23]	Outdoor	Person	Partially solved
	Single Gaussian [24]	Outdoor	Vehicle	Partially solved
	GMM + KDE [28]	Outdoor	Vehicle	solved
	GMM + Wavelet Transform [36]	Outdoor	Vehicle	solved
	PBAS [18]	Outdoor	Person	solved
Dynamic environment	MOG [26]	Indoor	Person	Partially solved
	Codebook [31]	Outdoor	Vehicle	solved
	PBAS [18]	Outdoor	Human	Partially solved
Shadow detection and Video Noise	Codebook [31]	Outdoor	Person	solved
	ViBe [33]	Outdoor	Car	solved

in [41] with pairwise region context information, which helps in increasing the shadow detection rate. Traditional methods do not work well in cluttered scenes and illumination conditions because many of them use hand-crafted features.

To overcome this problem, a proposed method [19] devises a shadow detection and removal technique for moving objects but it fails due to ghosting area and heavy noise. Another color invariant and geometric features-based method is discussed to detect shadow as a multi-stage process. The assumption is made that the shadow area is darkened than the foreground object [17]. After that, the result is checked using color invariance and shadow location concerning the object. But this method fails for detecting shadows with lower dark regions. The method proposed in [42] uses hue, saturation, and value (HSV) color space shadow properties to distinguish shadow pixels from moving objects. These features show that in the brightness component, cast shadows conceal the backdrop, whereas the saturation and hue spaces fluctuate within certain boundaries. The LAB color space method is introduced [43] which uses brightness values to detect shadows and observe that B-channel values (yellow to blue ratio) are lower in shadow regions than in non-shadow regions. Because it identifies each pixel as a shadow or non-shadow region without including nearby pixels, this approach produces erroneous labeling. We have categorized shadow detection and removal methods based on intensity and chromaticity features as shown in Table 2.

With the advancement of deep learning in recent years, approaches for processing moving object detection using Convolutional Neural Networks (CNN) have begun to emerge [44, 45]. Deep learning-based methods beat sample-based background modeling methods in most cases. But motion detection approaches require a large amount of labeled training sample data to train neural networks. Furthermore, CNN training necessitates some hardware assistance as well as a significant amount of time. On the other hand, sample-based background modeling methods do not face such issues. For this reason, we have not compared the proposed method with CNN-based methods in the experimental section due to their inflexibility.

Table 2 Classification of shadow detection and removal methods based on different types of shadow

Features	Approaches	Shadow Type	
		Cast Shadow	Self Shadow
Intensity based	Graph-based [40]	✓	
	MRF [41]	✓	
Chromaticity based	Color-based [19]	✓	
	HSV color space [42]	✓	✓
	LAB color space [43]		✓

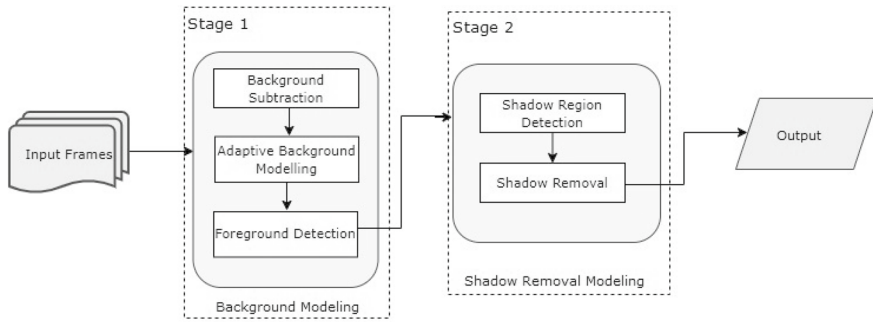


Fig. 3 Overview of the proposed framework

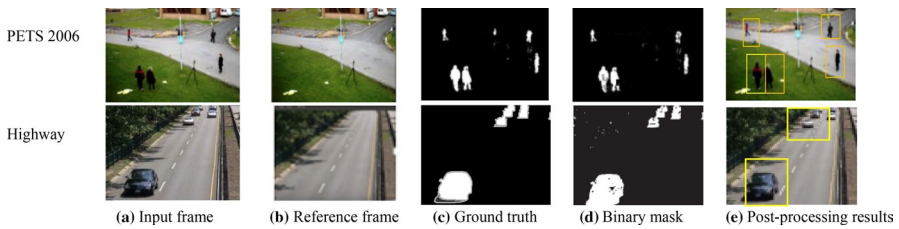


Fig. 4 Qualitative results under the category of background subtraction without shadow. See the results from left to right

3 Proposed method

This section discusses a detailed description of the proposed framework. The overview of the proposed system is depicted in Fig. 3, which consists of two stages: (i) background modeling and (ii) shadow removal modeling. The background modeling stage generates an adaptive threshold value for each frame pixel-wise to extract the binary mask for object segmentation. We take grey-level input frames for stage 1, in which a reference frame is generated from the initial few frames using the background subtraction method. Then, adaptive background modeling is applied to generate the automatic threshold value to update the background model for foreground detection. The visual representation of the first stage result is shown in Fig. 4. Sometimes, the foreground pixels are misclassified due to the existence of the shadow. The second step of the proposed framework is used to first find the shadow area and then get rid of the shadow pixels so that the pixels can be correctly classified which helps in accurate object detection. The visual representation of the second stage result is shown in Fig. 5. The second stage of the proposed framework is avoided if there is no shadow area in the result of the segmented objects.

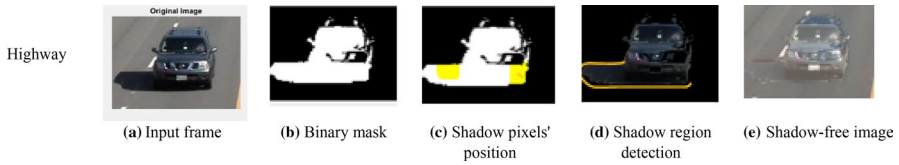


Fig. 5 Qualitative results of the Highway dataset under the category of background subtraction with shadow. See the results from left to right

3.1 Proposed adaptive background subtraction approach

The basic concept of the background subtraction model is to generate a reference frame called a background frame (consider a static background) and compare it with a new frame to classify each pixel belonging to the foreground or background by setting the manual threshold value which is a tedious task. Therefore, we propose a pixel-level background subtraction model using a two-level adaptive thresholding approach to segment the objects in a static background environment. Our key innovation is that a two-level per-pixel thresholds approach is used to generate the initial threshold for creating a background frame and then dynamically update the threshold value for each pixel depending on an estimated background frame. The proposed work is divided into three phases:

Phase 1: Initialize the reference frame for the background subtraction model.

Phase 2: Update the background subtraction model using adaptive background modeling.

Phase 3: Enhancement of the adaptive background subtraction model for foreground detection.

3.1.1 Phase 1: Initialize the reference frame for the background subtraction model

This phase is useful to generate a background frame called a reference frame from the initial few frames in consideration of not having moving objects in those frames. In this research work, we adapt the concept of the pixel-based averaging method [18] to generate the background frame by automatically calculating the initial threshold value. Here, we consider $B(x)$ initial frames $\{B_1(x), B_2(x), \dots, B_{n-1}(x), B_n(x)\}$ where x represents the location of pixels in the frame. We calculate the Mean ($M_1(x)$) by using Eq. (1) and Variance ($V_1(x)$) by using Eq. (2) of first frame to understand the statistical behavior of each pixel in the frame.

$$M_1(x) = \frac{1}{n} \sum_1^n B_1(x) \quad (1)$$

$$V_1(x) = (M_1(x) * M_1(x)) \quad (2)$$

An initial threshold is computed by using Eq. (3) to generate the reference frame. This step is used for the pixels' segmentation to generate the background frame.

$$\text{Threshold} = \sqrt{V_k(x) - (M_k(x) * M_k(x))} \quad (3)$$

The entire process of creating the initial background model for initial frames is summarized in Algorithm 1.

Algorithm 1 Computing Initial Threshold Value

Input: Consider B no of initial frames $\{B_1(x), B_2(x) \dots B_{n-1}(x), B_n(x)\}$

Output: Compute the initial background model and the threshold value

- An initial background model BGM frame, i.e., $M_k(x)$
- An initial Threshold value for pixel classification

Procedure:

- 1: Read one frame at a time from the initial frames
 - 2: Convert into a grey color
 - 3: Compute the mean value of the first frame
 $M_1(x) \leftarrow B_1(x)$
 - 4: Compute the variance of the first frame
 $V_1(x) \leftarrow M_1(x) * M_1(x)$
 - 5: Development of the background model
 for $k \leftarrow 2$ to T
 $M_k(x) \leftarrow M_{k-1}(x) + B_k(x)$
 $V_k(x) \leftarrow V_{k-1}(x) + M_{k-1}(x) * M_{k-1}(x)$
 end for
 - 6: Computation of initial threshold
 $\text{Threshold} \leftarrow \sqrt{\text{abs}(V_k(x) - M_k(x) * M_k(x))}$
-

3.1.2 Phase 2: Update the background subtraction model using adaptive background modeling

After initializing the background frame, the next step is to extract the moving objects by updating the background subtraction model for the remaining F frames $\{F_1(x), F_2(x), \dots, F_{n-1}(x), F_n(x)\}$. Here, moving objects of the first test frame $F_1(x)$ is segmented using the initial threshold value which is calculated from Algorithm 1. Then, to keep updating the threshold value adaptively for each frame, we propose a frame difference-based variance calculation method to calculate the threshold value for each frame at the pixel level. The background model is updated by generating a new threshold value T_n using the variance value of the test frame. This method helps in understanding the range of pixels' values for correct foreground and background segmentation, and also allows for gradual changes in the background frame. The frame differencing method is applied in testing frames using adaptive thresholding T_n to solve an issue of manual selection of threshold value for automatic segmentation of moving objects in a static scene. The frame differencing method for

pixel segmentation is shown in Eq. (4) and Eq. (5) where $f_n(x)$ is a current frame, $BGM_n(x)$ is a reference frame and T_n represents adaptive threshold value. If the pixel difference value of the current frame and reference frame is greater than the threshold value, then the pixel is classified as a motion pixel (foreground) otherwise pixel is classified as a static pixel (background).

$$\text{diff}_n(x) = |f_n(x) - BGM_n(x)| \quad (4)$$

$$\text{diff}_n(x) > T_n \quad (5)$$

The entire procedure of segmentation of moving objects for further frames using adaptive thresholding is shown in Algorithm 2.

Algorithm 2 *Updation process of background subtraction model*

Input: Consider N no of test frames $\{F_1(x), F_2(x)...F_{N-1}(x), F_N(x)\}$

Output: Compute background model and threshold

- Classify the state of each pixel as motion foreground or stationary background
- An updated Threshold value for pixel classification with adaptive learning

Procedure:

```

1: Read one frame at a time from test frames
2: Convert all frames into a grey color
3: Foreground Modelling and Extraction
   for k ← 1 to N
     if(n == 1)                                     // for the initial frame
       BGM1(x) ← Mk(x)                             // background model frame w.r.t.
       V1(x) ← Vk(x)                               initial mean, variance, and threshold
       (Threshold)n ← Threshold                     value
     else
       Vn(x) ← Vn-1(x)                             // for other frame
       BGMn(x) ← BGMn-1(x)
3.1 Initialise a frame difference matrices
     diffn(x) ← |fn(x) - BGMn(x)|                 // difference of respective pixels from
                                                         current frame and BGM frame
3.2 Foreground and background pixel classification
     If (diffn(x) ≥ Thresholdn)
       FGn(x) ← true                               // pixel is inmotion,i.e. partof the foreground
     else
       BGn(x) ← false                               // pixel is static, i.e. part of the background
3.3 Identify the boundary of extracted foreground
     Fboundary(x) ← (fn(x) * FGn(x) + BGMn(x) * BGn(x))
3.4 Update the background model
     BGMn+1(x) ← BGMn(x) + Fboundary(x)
     Vn(x) ← Vn-1(x) * Fboundary(x)2
     BGMn+1(x) ← BGMn(x)/n
     Vn+1(x) ← Vn-1(x)/n
3.5 Update the threshold value
     (Threshold)n+1 ← √abs(Vn+1(x) - BGMn+1(x) * BGMn+1(x))

```

3.1.3 Phase 3: Enhancement of the adaptive background subtraction model for foreground detection

The next step is to enhance the proposed adaptive background subtraction model using morphology operations with eight connected components. When the background model is updated, a few pixels are misclassified which causes an issue with holes in the inner region of the identified objects. To overcome this problem, we

incorporate morphology operations with the connected component method in the proposed background subtraction method. The connected component method helps in a grouping of similar kinds of pixels based on their connectivity and morphology operations are used to fill the gap in the internal region of the detected object. The morphological operations work as a filter on a binary mask generated from background modeling. Here, we use rectangular structure element $S(i, j) \in R$ i.e. object region of size (3×3) to (5×5) which is defined as a 2D function instead of a point set in binary imaging. The size of the structure element is dependent on frame resolution. We apply closed morphology operation which dilates the segmented image to fill small gaps in the interior region of the binary image and then erodes the dilated image to restore the original shape of the segmented frame. The closing is the dual operation $(A \cdot S)$ of the opening function and it is represented mathematically as given below [46]:

$$A \cdot S = \{A \oplus S\} \ominus S \quad (6)$$

$$(A \oplus S)(u, v) = \max_{(i,j \in S)} \{A(u + i, v + j) + S(i, j)\} \quad (7)$$

$$(A \ominus S)(u, v) = \min_{(i,j \in S)} \{A(u + i, v + j) - S(i, j)\} \quad (8)$$

where $(A \oplus S)(u, v)$ represents dilation operation and $(A \ominus S)(u, v)$ represents erosion operation.

After applying morphology operations, we use the eight-way pixel connectivity connected component method, which group the pixels that have similar intensity. This method helps restore the original image with detected objects as foreground. This operation is useful to detect objects with boundary boxes that may help in identifying the location of the objects. Though our background subtraction model is enhanced with post-processing operations but some false positive pixels are also detected due to the occurrence of shadow pixels, which degrade the performance of the model.

3.2 Proposed shadow region detection and removal model

The removal of shadows has become a key step to improve the performance of the surveillance system. A shadow is considered a dark region and is generated by an opaque object blocking light from a light source. The presence of shadow disturbs the characteristics of a segmented object by incorrectly classifying the foreground pixels. When a shadow is cast by static objects (like a building, tree, pillar, parked vehicle, etc.) in such cases, methods to detect moving objects using background subtraction do not suffer from the static shadow because these types of shadows are considered as background. But if shadow appears in a moving object (foreground part) then false positives may increase due to the misclassification of pixels. Our proposed framework solves the concern issue by integrating shadow region detection and removal steps with the background subtraction model. Our method was inspired by [19] and proposed a chromaticity-based shadow region detector to detect

the shadow pixels. Also, an intensity ratio-based YCbCr color space model has been proposed to remove the shadow pixels. The model works with a generated binary mask using a background subtraction model and color frame of foreground to check the existence of shadow. This second stage of our proposed framework consists of two phases:

- Phase 1: Detection of the shadow region
- Phase 2: Removal of shadow

3.2.1 Phase 1: Detection of the shadow region

The shadow region detection is a primary step in shadow removal modeling with an assumption that pixels under the shadow region are darker than the background region. However, a good shadow region predictor model requires predicting the range of pixel intensity that lies under the shadow region. In this research work, a chromaticity-based shadow region detection approach is adopted to know the statistics of pixels intensity as shadow region (S) which has a lower intensity value than the non-shadow region (\tilde{S}) depends on illumination conditions. Due to illumination conditions, the intensity of the shadow region can be sharp or smooth. To solve this issue, we propose an intensity ratio-based YCbCr color correction model in which luminance information is stored by the Y channel and color information is stored by two other color channels Cb and Cr. Our proposed method takes the color input frame and binary mask obtained from stage 1 and processes those frames to get the shadow region if exists with an assumption that pixels are linearly related to the intensity value of incident light. To start searching for shadow region, first, convert the color frame into YCbCr color space and compute the summation of the average pixel intensity ratio $R_s(x)$ of each YCbCr channel between test frames $F_n(x)$ and background model frame $BGM_n(x)$ as shown in Eq. (9)

$$R_s(x) = \frac{F_n(x)}{BGM_n(x)} \tag{9}$$

The next step is to identify pixels under the shadow region by scanning the frame horizontally ($H_n|i|$) and vertically ($V_n|j|$). Then, we count the total number of shadow pixels (N_s) in shadow region as formulated by Eq. (10) where $j_{n.min}$ and $j_{n.max}$ represent minimum and maximum pixels value at the vertical level. While $i_{n.min}$ and $i_{n.max}$ represents minimum and maximum pixels value at the horizontal level.

$$N_s = \sum_{j=j_{n.min}}^{j_{n.min}+j_{n.max}} V_n|j| + \sum_{i=i_{n.min}}^{i_{n.min}+i_{n.max}} H_n|i| \tag{10}$$

where $j \in \{0, \text{Width} - 1\}$ and $i \in \{0, \text{Height} - 1\}$.

We also compute mean $Y_{\text{Mean}}(x)$ and standard deviation $Y_{\text{SD}}(x)$ of other test frames $F_n(x)$ as formulated by Eq. (11) and Eq. (12), respectively, to classify the shadow pixels. Here, the mean value works as the intensity threshold value

if pixel intensity is less than the standard deviation than pixels come in shadow region (S) otherwise it is considered as non-shadow region (\tilde{S}).

$$Y_{\text{Mean}}(x) = \frac{1}{N} \sum_{i=1}^n I_i(x) \quad (11)$$

$$Y_{\text{SD}}(x) = \sqrt{\frac{1}{N} \sum_{i=1}^n (I_i(x) - M(x))^2} \quad (12)$$

It is observed that a shadow contour is usually associated with a moving object, so it is required to search adjacent pixels to find the complete shadow region as shown in Eq. (13) and Eq. (14), respectively.

$$\begin{aligned} &\text{if}(H_n|i| \neq 0 \ \&\& \ H_n|i + Y_{\text{Mean}}(x)| < N_s) \\ i_{\text{HAdjacent}} = i : &\text{ where } 0 \leq i \leq \text{height} \end{aligned} \quad (13)$$

$$\begin{aligned} &\text{if}(V_n|j| \neq 0 \ \&\& \ V_n|j + Y_{\text{Mean}}(x)| < N_s) \\ j_{\text{VAdjacent}} = j : &\text{ where } 0 \leq j \leq \text{width} \end{aligned} \quad (14)$$

The entire procedure to detect shadow region as shown in Algorithm 3.

Algorithm 3: Shadow region detection

Input: RGB frame and generated binary mask from proposed background subtraction model

Output: Shadow region in the frame

Procedure:

1: Read the RGB frame and convert it into YCbCr color space

2: Compute the average pixel intensity ratio of each YCbCr color channel frame

$$R_s(x) \leftarrow \frac{F_n(x)}{BGM_n(x)} \quad // \text{ Refer to Eq.(9)}$$

3: Compute mean (Y_{Mean}) and standard deviation (Y_{SD}) of Y color channel

4: Search the shadow region window and calculate the total number of shadow pixels

$$S_n \leftarrow V_s(x) + H_s(x) \quad // \text{ Refer Eq.(10)}$$

5: Classify shadow and non-shadow region

$$\text{If } (YCbCr(i, j) < Y_{\text{Mean}} - \frac{Y_{\text{SD}}}{3})$$

$$S(x) \leftarrow F_n \quad // F_n(x) \text{ is considered as shadow pixel}$$

$$\text{else if } (YCbCr(i, j) > Y_{\text{Mean}} - \frac{Y_{\text{SD}}}{3})$$

$$\tilde{S}(x) \leftarrow F_n \quad // F_n(x) \text{ is considered as non-shadow pixel}$$

3.2.2 Phase 2: Removal of shadow

The next step of our approach is to remove shadow pixels from the processed frame. Here, we use the YCbCr color correction method to correct the luminance property generated by the Y channel while the light model-based method is used to correct the color properties generated by Cb and Cr channels. The intensity $I_i(x)$ of shadow removal method can be represented by Eq. (15)

$$I_i(x) = \left(t_i \cos \theta_i \widetilde{S}_i(x) + S_i(x) \right) R_i \quad (15)$$

where $\widetilde{S}_i(x)$ represents non-shadow pixels, $S_i(x)$ represents shadow pixels, R_i represents surface reflectance and θ_i represents an angle of light on the object and t_i represent direct light attenuation factor where if $t_i = 1$ then the object is in the non-shadow region.

In the shadow removal method, we compute an average of non-shadow $\widetilde{S}_{avg}(x)$ and shadow $S_{avg}(x)$ pixels intensity and then calculate a ratio between averaging of non-shadow and shadow pixels are computed by Eq. (16)

$$r = \widetilde{S}_{avg}(x) / S_{avg}(x) \quad (16)$$

To identify the difference between shadow and non-shadow regions, calculate the difference in average values of shadow and non-shadow pixel intensity as illustrated in Eq. (17)

$$\text{diff}_s = \widetilde{S}_{avg}(x) - S_{avg}(x) \quad (17)$$

The ratio of two pixels' values is not the same in all Y, Cb, and Cr color channels so color correction is required for the proper removal of shadow pixels. It is required to add the difference and ratio values in the Y, Cb, and Cr color channels and then converted them to RGB channels. We compare the pixels' value with standard deviation as the threshold value generated by Eq. (12), once the pixel value is less than the threshold value then remove that pixel by associating the value 0 which is considered as part of the background. Finally, a closed morphology operation is applied to fill the gaps inside the detected region followed by connected components to detect the shadow-free foreground objects. Algorithm 4 illustrates the proposed shadow removal approach.

Algorithm 4 Shadow Removal**Input:** Shadow region frame**Output:** Shadow-free frame**Procedure:**

- 1: Compute the average shadow $S_{avg}(x)$ and non-shadow $\widetilde{S_{avg}}(x)$ pixels intensity
- 2: Calculate the difference between both average values

$$diff_s \leftarrow S_{avg}(x) - \widetilde{S_{avg}}(x)$$

- 3: Compute Ratio of average pixels calculated from step 1

$$r \leftarrow \frac{S_{avg}(x)}{\widetilde{S_{avg}}(x)}$$

- 4: Add ratio and difference values to each color channel for the segmentation of the shadow region

$$Y(x) \leftarrow r + diff_s$$

$$Cb(x) \leftarrow r + diff_s$$

$$Cr(x) \leftarrow r + diff_s$$

- 5: Apply closed morphology operations to fill interior gaps
- 6: Convert YCbCr into RGB color space

4 Experiments and result discussion

In literature [9, 39, 47] evaluation of the background subtraction methods are based on human perception (subjective evaluation). It is difficult to assess the image quality based on subjective evaluation because it depends on capturing device, distance, viewing angle, environment, and human perception. This type of evaluation is time-consuming and impractical in a real-world scenario. Therefore, it is needed to use statistical evaluation to evaluate the performance of the methods accurately. The efficiency of the proposed method is experimenting on benchmark datasets and comparing the performance of the proposed method with state-of-the-art methods using standard performance measures.

4.1 Datasets description

The proposed work is evaluated on three widely used datasets (i) Highway [12] (ii) PETS 2006 [12] and (iii) SBU [48] as shown in Table 3. The reason to consider these datasets on accountability of motion or change detection in an outdoor and indoor scene. These datasets are available in video as well as in frame format. Few frames of the Highway dataset deal with background motion noise and mild shadow. In PETS 2006 dataset, few frames deal with illumination changes but no shadow. SBU dataset completely deals with the shadow category. Frame resizing (512×512) has been done for all frames of the PETS 2006 and SBU datasets for

Table 3 Description of the benchmark datasets

Datasets	Total frames	Frame size	Ground Truth	Scene	Shadow size
PETS 2006	1200	720×576	1200	Indoor and outdoor	No shadow
Highway	1700	320×240	100	Outdoor vehicle	Medium
SBU	4085	512×512	4085	Indoor and outdoor	Medium and large

better evaluation. Ground truth is associated with each dataset for measuring the performance of the proposed method with state-of-the-art methods.

4.1.1 Ground truth labels

The benchmark datasets are available with ground truth labeled data for comparing the robustness and efficiency of the proposed method with existing methods. Each frame in available datasets is annotated manually at the pixel level as shown in Table 4.

4.2 Effectiveness of the proposed model

This research is based on the principle in which object segmentation is done when it appears the first time in the frame so the background model needs to be updated to generate segmentation results. In this section, experimental results are shown qualitatively (visual perception) and quantitatively based on the available ground truth frames to prove the effectiveness of the proposed method.

4.2.1 Qualitative results

Our approach is based on two-level adaptive thresholding with chromaticity-based YCbCr color correction-based shadow removal method to generate a binary mask for foreground segmentation and detect the shadow-free foreground. This

Table 4 Ground truth labeling at pixel level

Sr No	Ground Truth Labeling	Gray Scale Pixel Value
1.	Static	Gray scale pixel value 0 is considered a static or background scene
2.	Shadow	Pixels ranging between 50 to 70 grayscale values are labeled with shadow region
3.	Non-ROI	The first few 100–200 frames of each dataset are assigned as NON-ROI frames to avoid errors in evaluation metrics
4.	Moving	Moving pixels are labeled with gray level scale value 255 for foreground detection

section covers the experimental analysis of each dataset under two categories (i) background subtraction without shadow and (ii) background subtraction with shadow. The initial value of the threshold is experimentally computed and lies between 40 to 50 grayscale pixel range for generating a reference frame. Once the initial threshold value is computed then foreground segmentation is done based on a threshold value for correct pixel classification whether pixels are belonging to the foreground or background. After foreground segmentation, post-processing operations like closed morphology using rectangular structure element (SE) of size (3×3) for the Highway dataset and (5×5) for the PETS 2006 dataset are applied to fill the gap in the internal region of the segmented results. Then, blob analysis using 8 connected components is applied to obtain the detected objects with a boundary box. The size of the structural elements is experimented tested. After experimentation, it is found that a larger value of structure elements deforms the shape of the blob, due to the larger shape of a structure element, multiple blobs are connected into a single one.

The qualitative results of the proposed background subtraction model with post-processing operations are shown in Fig. 4, 5, and 6, respectively, under the categories of background subtraction results without and with shadow. We considered the first 150 frames of the PETS 2006 dataset and the first 100 frames of the Highway dataset for generating the reference frame, as shown in Fig. 4b. Here, the visual results of the 551st input frame of the PETS 2006 dataset and the 692nd input frame of the Highway dataset, along with their ground truth frames, have been shown, respectively.

However, most background subtraction methods [8, 9, 13] can correctly classify the pixels but generate more false positives when shadow exists in a segmented frame. Though shadow is considered part of the background in the frame sometimes it can be misclassified as foreground if the background color and foreground color are nearly similar (refer to Fig. 5 of the Highway dataset). Few frames in the Highway dataset suffer from a mild shadow, so the second stage of the proposed method works well when a shadow region is detected while processing the frames. The proposed method also has the benefit of detecting and eliminating shadows in the processed frame to decrease false positives. Figure 5 displays the visual results of the 387th input frame of the Highway dataset under the category of background subtraction with shadow.

To test the ability of the proposed method, we consider medium and large sizes of shadows, which should be associated with the required foreground objects. The

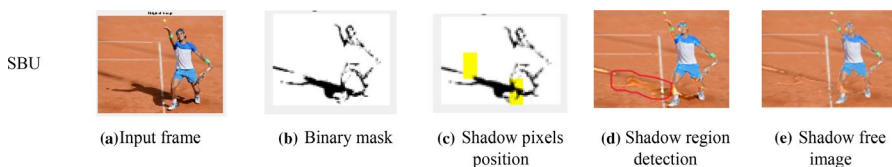


Fig. 6 Qualitative results of SBU dataset under the category of background subtraction with shadow. See the result from left to right

experiments are conducted on the SBU dataset, in which every image is associated with shadow pixels. The quantitative results of the 25th frame of the SBU dataset under the category of background subtraction with shadow are shown in Fig. 6.

The proposed method also detects and removes single and multiple shadows from the captured scene. Though the SBU dataset is designed specifically for shadow detection and removal, the second stage of our method, i.e., shadow detection and removal, is tested on various images of the said dataset, and it is observed that the proposed method also works well if only the shadow is visible with no foreground object. Visual representations of the same are shown in Fig. 7 (consider the results row-wise) where the first row represents the qualitative result of 42th image sequence when only a single shadow is available and the second row represents the qualitative results of 135th image sequence when multiple shadows are available.

After experimentation, it was observed that some internal holes were generated in the shadow removal area, which could be recovered by closed morphology operations. In this paper, we use structure elements (3×3) and (5×5) during morphology operations to generate smooth results.

4.2.2 Comparisons of qualitative results of the proposed background subtraction method with state-of-the-art-methods

The foreground detection and shadow removal become challenging due to illumination changes and similarity in the foreground and background color. The comparison has been done based on two stages model. The first stage is used to generate a binary mask using a background subtraction model. The performance has been compared with well-known methods which work at the pixel level (i) Temporal Average Filter [8], (ii) Running Average Gaussian (RAG) [26], (iii) Adaptive GMM [27], (iv) WeP-BAS [34], (v) ViBe [33]. These methods are chosen on the requirement of a mapping threshold value for foreground segmentation. While the Second stage of the proposed model is used to detect and remove shadow if exists and the result is compared with four well-known methods (i) LAB Color method [43], (ii) HSV Color

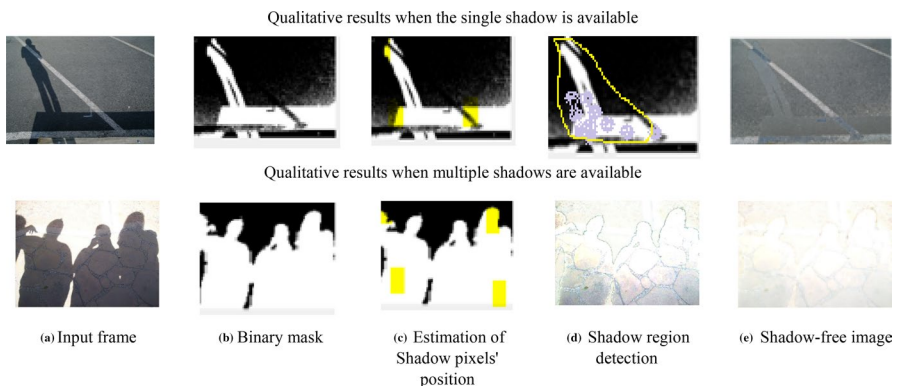


Fig. 7 Qualitative results of SBU dataset when single and multiple shadows are available. See the results row-wise

method [42], (iii) Sample based method [47] and (iv) Color Constancy method [19]. These methods are chosen to fulfill the requirement of shadow color comparison at a pixel level. The visual representation of the Highway dataset for various frames is shown in Fig. 8 where cars traveling on the highway should be classified as foreground (FG), whereas the rest of the scene, including swaying tree branches beside the roadway, should be classified as background (BG). After experimentation, it has been observed that the adaptive threshold value lies between 87 to 110-pixel values to segment generate the binary mask of foreground objects. The proposed reference frame updated mechanism helps in identifying the movement of the objects for better binary segmentation.

The visual results of the PETS 2006 dataset for outdoor video sequences are shown in Fig. 9 where the adaptive threshold value has been automatically calculated that range lies between 90 to 125 pixels values. As per results, pedestrians and bicyclists are passing on the street where illumination changes occur in frames. The proposed algorithm segments pedestrians and bicyclists more accurately than other methods. However, ViBe [33] achieves good results in the classification of the foreground but some noise is associated when sudden illumination changes happen as shown in the frame no. 942. Our method is enriched with post-proceeding operations to reduce the noise in the segmented frame.

4.2.3 Comparisons of qualitative results of proposed shadow detection and removal method with state-of-the-art-methods

The SBU dataset contains varieties of images that are associated with shadow. The visual representation of the SBU dataset for shadow detection and removal is shown in Fig. 10. Our method generates the results of shadow removal of color images that are associated with binary masks. The proposed method is tested on three test cases (I) when an object is associated with shadow, (II) only shadow is available, and (III) multiple shadows are available. Our method presents smooth results than

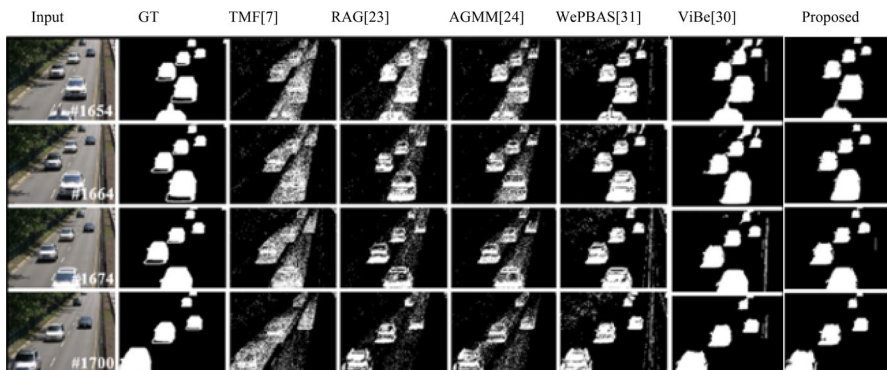


Fig. 8 Comparisons of background subtraction of different frames on the Highway dataset. See the results column-wise from left to right: (i) Input frame, (ii) GT=Ground Truth, (iii)TMF=Temporal Medium Filter, (iv) RAG=Running Average Gaussian, (v)AGMM= Adaptive Gaussian Mixture Model, (vi) WePBAS= Weighted Pixel based Adaptive Segmenter,(vii) ViBe and (viii) Proposed method

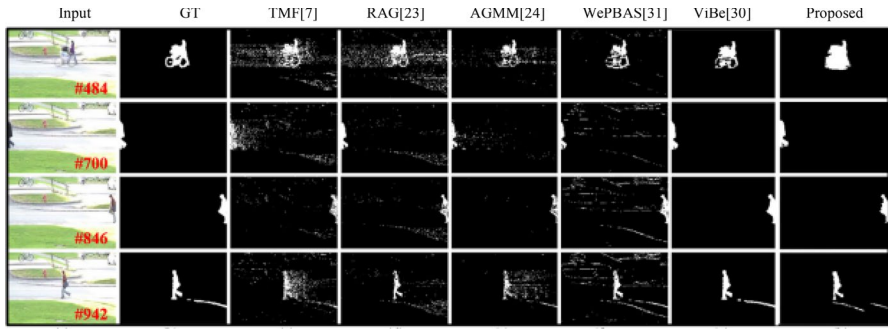


Fig. 9 Comparisons of background subtraction of different frames on PETS 2006 dataset. See the results column-wise from left to right: (i) Input frame, (ii) GT=Ground Truth, (iii) TMF=Temporal Medium Filter, (iv) RAG=Running Average Gaussian, (v)AGMM= Adaptive Gaussian Mixture Model, (vi) WePBAS= Weighted Pixel based Adaptive Segmenter,(vii) ViBe and (viii) Proposed method

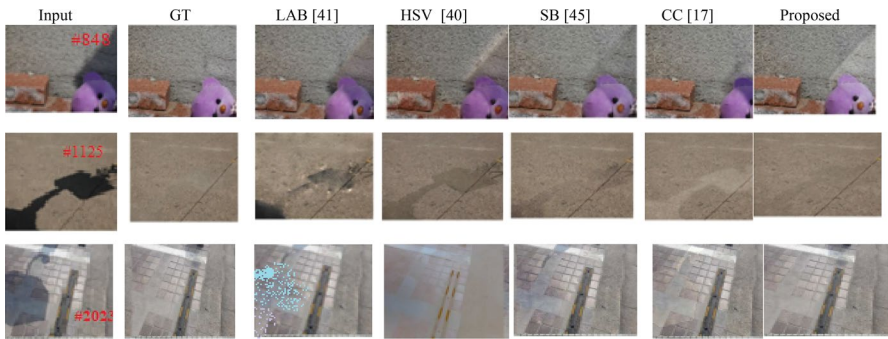


Fig. 10 Comparisons of shadow detection and removal part of different frames on SBU dataset. See the results column-wise from left to right: (i) Input frame, (ii) GT=Ground Truth, (iii) LAB Color Space method (iv) HSV Color Space method, (v) SB=Sample based method, (vi) CC=Color constancy method, and (vii) Proposed

state-of-the-art methods due to the value of the intensity ratio of the color channel. The advantage of our proposed model is to remove single and multiple shadows. There is one limitation of the proposed shadow detection and removal stage is that it is not able to remove shadow smoothly if the shadow is non-uniform as shown in image sequence 1023(third row in Fig. 10).

4.2.4 Quantitative Results

This section discusses a quantitative analysis of the proposed work. The experiments are conducted on three benchmark datasets (as discussed in Sect. 4.1) with available ground truth data to test the efficiency of the proposed work. The correctness of the proposed background subtraction model is measured by six statistical measures *Precision* considers the correct segmentation of pixels, *Recall* represents foreground pixels are correctly detected as foreground available in the ground truth dataset, *false negative*

rate (*FNR*) considers the proportion of foreground pixels are incorrectly classified as background, *false positive rate (FPR)* considers the proportion of background pixels are incorrectly classified as foreground, *F-Measure* is a weighted harmonic mean of precision and recall and lastly *percentage of correct classification (PCC)* is most important evaluation metrics for correct pixels classification. All of these measures are sensitive to the reference frame. The performance measures are stated as [47]:

$$\text{Recall} = \frac{tp}{tp + fn} \quad (18)$$

$$\text{Precision} = \frac{tp}{tp + fp} \quad (19)$$

$$\text{FPR} = \frac{fp}{fp + tn} \quad (20)$$

$$\text{FNR} = \frac{fn}{tp + fn} \quad (21)$$

$$F - \text{Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (22)$$

$$\text{PCC} = \frac{tp + tn}{tp + fp + tn + fn} \quad (23)$$

Here tp =true positive considering foreground pixels are correctly detected, tn =true negative considering background pixels are correctly detected, fp =false positive considering pixels falsely detected as foreground, and fn =false negative considering pixels falsely detected as background.

The presented work detects shadow pixels and their removal from extracted frames to achieve good accuracy. Therefore, we use another two evaluation measures (i) shadow detection rate (η) which is related to correctly identified shadow pixels and (ii) shadow discrimination rate (ξ) which is used to remove the shadow pixels to identify the difference between foreground objects and shadow area. The equations of these parameters are given below [43]:

$$\eta = \frac{tp_s}{tp_s + fn_s} \times 100 \quad (24)$$

$$\xi = \frac{\overline{tp_f}}{tp_f + fn_f} \times 100 \quad (25)$$

Here tp_s represents shadow pixels are correctly classified, fn_s represents foreground pixels are incorrectly classified as a shadow, tp_f represents correctly

Table 5 Statistical evaluation of the proposed method with existing methods on the PETS 2006 dataset

Methods	Precision	Recall	FPR	FNR	F-Measure	PCC
TMF [8]	0.5127	0.5284	0.3261	0.1002	0.5149	0.6872
RAG [26]	0.5893	0.6255	0.3171	0.1142	0.5993	0.7436
AGMM [27]	0.8303	0.7282	0.1615	0.0662	0.7934	0.8364
WePBAS [34]	0.8481	0.7615	0.1063	0.0216	0.8195	0.9053
ViBe [33]	0.8790	0.8191	0.0657	0.0153	0.8654	0.9342
Proposed	0.9496	0.8716	0.0341	0.0068	0.9449	0.9646

Bold values depict the best result

Table 6 Statistical evaluation of the proposed method with existing methods on the Highway dataset

Methods	Precision	Recall	FPR	FNR	F-Measure	PCC
TMF [8]	0.5024	0.5358	0.3083	0.1045	0.5145	0.6683
RAG [26]	0.5762	0.6057	0.3054	0.1501	0.5846	0.7162
AGMM [27]	0.8182	0.6948	0.2957	0.1654	0.7452	0.7296
WePBAS [34]	0.8250	0.7661	0.1092	0.1203	0.8047	0.8992
ViBe [33]	0.8590	0.7525	0.0765	0.0788	0.8542	0.9221
Proposed	0.9254	0.9041	0.0434	0.0743	0.9247	0.9573

Bold values depict the best result

identified foreground pixels, fn_f represents foreground pixels incorrectly classified as a shadow or background pixels, tp_f represents ground truth foreground pixels.

4.2.4.1 Performance analysis of the proposed background subtraction method with existing methods After experimentation, it has been observed that the standard deviation value which is used to calculate the initial threshold for background modeling lies between 4.4 and 5.1 for the Highway dataset, 5.3 to 5.9 for PETS 2006 dataset, and 4.8 to 6.4 for the SBU dataset. An initial threshold value is used to generate initial binary segmentation from the reference frame and current frame. After that, the reference frame has been updated by calculating the adaptive threshold value at a pixel level to generate the binary mask for foreground extraction. The efficiency of the proposed work has been compared with the state-of-the-art methods in terms of background modeling and shadow removal. Methods are compared based on selected parameter values as used in [34, 47]. Tables 5 and 6 represent a quantitative analysis of the proposed method after post-processing operations and observe that the proposed method can generate better results than compared methods.

The graphical representation of Table 5 (FPR and PCC values) is shown in Fig. 11 which demonstrates that the proposed method generates less FPR value and more PCC value for PETS 2006 dataset.

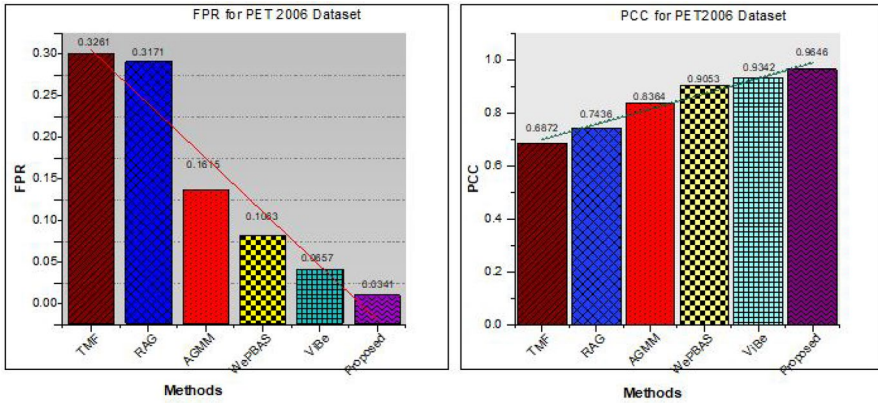


Fig. 11 Comparisons of FPR and PCC values of the proposed method with existing methods on the PETS 2006 dataset

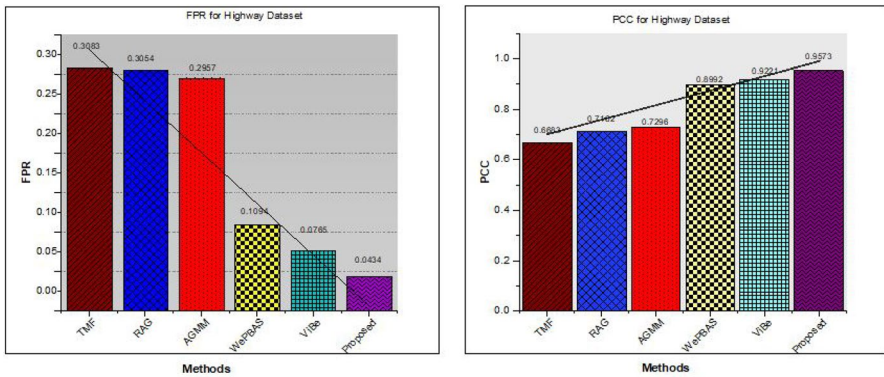


Fig. 12 Comparisons of FPR and PCC values of the proposed method with existing methods on the Highway dataset

The graphical representation of Table 6 (FPR and PCC values) is shown in Fig. 12 where less FPR values and more PCC values are considered for the proposed work.

According to the test results, the proposed method achieved the highest percentage of correct classification (PCC). It has been observed that the adaptive threshold value should not be very high, otherwise foreground objects will be considered background due to some similarity between the foreground and background scenes. However, WePBAS [34] and ViBe [33] are the most commonly used algorithms, but they do not perform well when shadow pixels are available in a high-contrast input frame. On the other hand, higher PCC values depend on lower FPR (false positive rates) values if pixels are correctly classified as foreground or background. The proposed method generates less FPR value and more PCC value for the PETS 2006 and Highway datasets, as shown in Figs. 11 and 12, respectively. TMF (Temporal

Median Filter) [8] and RAG (Running Average Gaussian) [26] produced more false positive values and less correct classification of pixels than the proposed method because of the effect of choosing a manual threshold value on their performance. There are few frames in the Highway dataset are associated with shadow, the Adaptive GMM [27], WePBAS [34], and ViBe [33] methods produced more pixel misclassifications (considering shadow pixels as foreground) and increased the false positive rate (FPR). It has been observed that in the Highway dataset, RAG [26] shows bad recall when more pixels are misclassified due to long shadows. The proposed method has the capability of detecting and removing shadow pixels, which generates less misclassification of pixels and a lower false-positive rate than other methods.

4.2.4.2 Performance analysis of the proposed shadow detection and removal method with existing methods

Furthermore, the proposed work is compared with existing methods in respective shadow detection and removal. The compared methods are chosen based on having the property to deal with shadows in the input frame. Performance comparisons are done using two performance measures such as shadow detection rate (η) and shadow discrimination rate (ξ). The methods are evaluated on two benchmark datasets consisting of shadow frames (i) Highway (as few frames are associated with shadows) and (ii) SBU dataset (all input images are associated with shadow) along with three test cases: (I) Shadow is associated with objects, (II) only shadow is available and (III) Multiple shadows are available. The PETS 2006 dataset is not evaluated as no shadow pixels are available.

In the Highway dataset, object visibility is a must to check the effectiveness of the system therefore only test case (I) is tested. We consider random 35 sample frames of the Highway dataset to test shadow detection and discrimination rate. The performance of the proposed method is illustrated in Fig. 13 where black lines indicate the shadow detection rate and red lines indicate the shadow discrimination rate. The average qualitative analysis for detected shadow pixels on sample frames is demonstrated in Table 7 where the increased shadow detection and discrimination rate could see as shadow pixels are properly detected and removed by the proposed method.

SBU dataset is associated with shadows and a qualitative comparison of shadow detection and shadow removal rate of the proposed method for all three test cases is shown in Fig. 14 where peaks represent higher values of detection and discrimination. The average qualitative evaluation of the proposed method is compared with other methods as shown in Table 8 for all three test cases. The shadow detection rate is dependent on the luminance color distance between shadow and non-shadow pixels. The compared methods mostly reject the shadow pixels when the luminance distance is very small. After experimentation, it is observed that the proposed method detects and removes shadows even if multiple shadows are available. However, it does not generate better results than the Color Constancy method [19] due to non-uniformity in shadow shape.

The shadow detection rate is dependent on the luminance color distance between shadow and non-shadow pixels. The compared methods mostly reject the shadow pixels when the luminance distance is very small. After experimentation, it has been

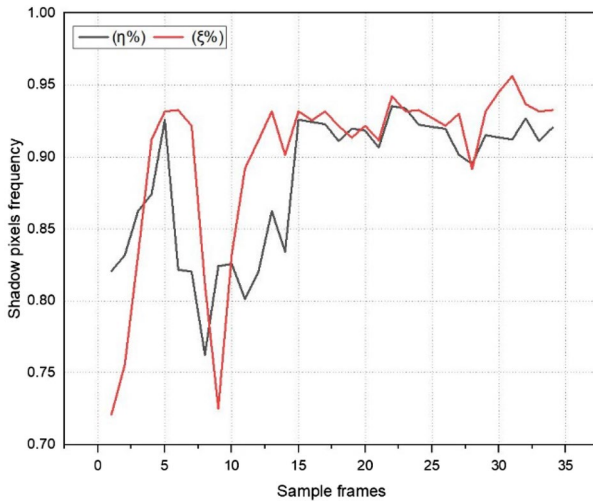
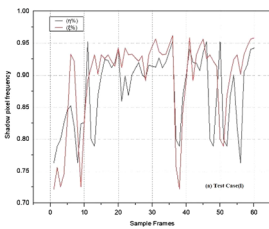


Fig. 13 Performance analysis of proposed shadow detection (η) and removal rate (ξ) of sample frames on the Highway dataset

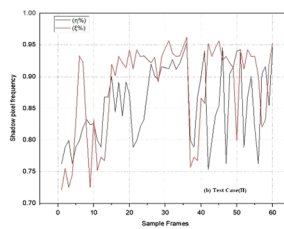
Table 7 Average quantitative analysis of proposed method with existing method on Highway dataset for the test case (I)

Methods	Shadow detection rate (η) in %	shadow discrimination rate (ξ) in %
LAB [43]	82.17%	74.11%
HSV [42]	82.02%	72.52%
Sample Based [47]	86.26%	83.23%
Color Constancy [19]	87.41%	91.21%
Proposed	92.57%	93.16%

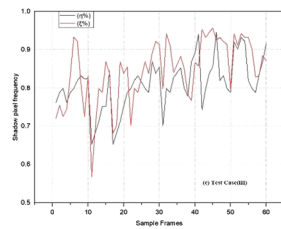
*Test case(I): Object is associated with shadow. Boldface values depict better results



(a) Test Case(I)



(b) Test Case(II)



(c) Test Case(III)

Fig. 14 Comparison of shadow detection (η) and removal rate (ξ) of the proposed method in the SBU dataset for all three test cases

Table 8 Average quantitative analysis of the proposed method with the existing method on the SBU dataset for test case (I), (II), and (III)

Test Cases	Methods	Shadow detection rate (η) in %	shadow discrimination rate (ξ) in %
Test Case (I)	LAB [43]	79.93%	75.68%
	HSV [42]	78.86%	72.20%
	Sample Based [47]	86.82%	84.43%
	Color Constancy [19]	90.03%	89.11%
	Proposed	94.04%	95.81%
Test Case (II)	LAB [43]	75.34%	77.28%
	HSV [42]	79.85%	76.75%
	Sample Based [47]	83.82%	86.62%
	Color Constancy [19]	85.46%	85.76%
	Proposed	94.57%	95.23%
Test Case (III)	LAB [43]	65.30%	56.78%
	HSV [42]	71.24%	69.26%
	Sample Based [47]	71.22%	67.88%
	Color Constancy [19]	89.32%	88.42%
	Proposed	91.04%	87.09%

*Test case(I): object is associated with shadow, Test case(II): only shadow is available, Test case(III): multiple shadows are available. Boldface values depict better results

observed that the proposed method detects and removes shadow pixels' when luminance changes. Our method can eliminate shadow pixels when multiple shadows are available but it does not generate a better result than the Color Constancy method [19] non-uniformity shadow shape has occurred (as shown in Fig. 10 last row).

4.2.5 Performance analysis

It has been observed that the proposed background subtraction model generates better foreground results by employing an adaptive thresholding procedure at the pixel level. The first stage of the model helps in identifying the initial threshold value to generate a reference frame, and then the background subtraction model is updated for each frame to segment the object from the scene. After some experimentation, it was found that most threshold values covered the pixels between 40 and 110. It is also observed that our proposed first stage of the method can segment the foreground object when it starts appearing for the first time in the frame. The quantitative and qualitative results show that the proposed model works well compared to state-of-the-art methods. Moreover, the second stage of the proposed method is also capable of detecting and removing shadow pixels from the input sequences, which helps reduce pixels' misclassifications. On a sunny day, a shadow can be considered a hard shadow if it is cast by an object. During experimentation, hard shadows are ignored because they are considered part of the background and do not increase false positive rates. After experimentation, it

was observed that the range of shadow pixels is between 30 and 70. Another significant finding is that the interior of the objects does not have shadow pixels, and any holes that do exist can be filled using post-processing techniques.

5 Conclusion

This research solves two important issues in moving object detection: (i) choosing an appropriate threshold value for foreground segmentation and (ii) misclassification of pixels due to shadow. Since background subtraction and shadow removal are typically discussed separately, our contribution is to solve the two to improve the performance of video surveillance systems for both indoor and outdoor activity recognition. The proposed ABGS segmenter generates a background subtraction method using a two-level adaptive thresholding approach to create a binary mask at the pixel level. This method is integrated with a chromaticity-based shadow removal method, which detects shadow pixels and removes them appropriately without distorting the shape of the object. The advantage of the proposed method is to remove the shadow pixels automatically even if a single shadow or multiple shadows are available in frames. The results of the experiments are evaluated on three benchmark datasets (Highway, PETS 2006, and SBU), which outperform state-of-the-art methods concerning six performance measures. We have also evaluated the performance of the shadow removal stage on three test cases with two statistical measures, considering the shadow detection rate (η) and shadow discrimination rate (ξ). After experimentation, it is observed that our proposed framework is robust and reliable over other methods. The only limitation is that it is not able to remove the shadow pixels due to non-uniformity in shadow shape casting on objects. We can extend our work to detect occluded objects to identify the correct shape in moving object detection which helps in improving the potential of the proposed method.

Acknowledgements Not applicable.

Authors' contributions The manuscript has been written by author 1 and reviewed by author 2.

Funding Not applicable.

Availability of data and materials Data are easily available online. http://changedetection.net/https://www3.cs.stonybrook.edu/~cvl/projects/shadow_noisy_label/index.html.

Declarations

Conflict of interests The authors declare that they have no competing interests.

Consent for publication Not applicable.

Ethics approval Not applicable.

consent to participate Not applicable.

References

1. Cheng FC, Huang SC, Ruan SJ (2010) "Advanced motion detection for intelligent video surveillance systems," in Proceedings of the ACM Symposium on Applied Computing, 983–984. <https://doi.org/10.1145/1774088.1774295>.
2. Bouwmans T (2014) Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*. <https://doi.org/10.1016/j.cosrev.2014.04.001>
3. Shaikh SH, Saeed K, Chaki N (2014) Moving Object Detection Using Background Subtraction. In: Shaikh SH, Saeed K, Chaki N (eds) *Moving Object Detection Using Background Subtraction*. Springer International Publishing, Cham, pp 15–23. https://doi.org/10.1007/978-3-319-07386-6_3
4. K. Sehairi, F. Chouireb, and J. Meunier (2015) "Comparison study between different automatic threshold algorithms for motion detection," In 4th Int. Conf. on Electrical Engineering (ICEE 2015), 1–8, <https://doi.org/10.1109/INTEE.2015.7416840>
5. Piccardi M, "Background subtraction techniques: A review," 2004. doi: <https://doi.org/10.1109/ICSMC.2004.1400815>
6. Garcia-Garcia B, Bouwmans T, Silva AJR (2020) Background subtraction in real applications: Challenges, current models and future directions. *Compt Sci Rev* 35:100204. <https://doi.org/10.1016/j.cosrev.2019.100204>
7. Sajid H, Cheung SS (2015) "Background subtraction for static & moving camera," In 2015 IEEE International Conference on Image Processing (ICIP). 4530–4534. <https://doi.org/10.1109/ICIP.2015.7351664>.
8. Zhang R, Ding J (2012) Object tracking and detecting based on adaptive background subtraction. *Proced Eng*. <https://doi.org/10.1016/j.proeng.2012.01.139>
9. Zivkovic Z (2004) "Improved adaptive Gaussian mixture model for background subtraction," In: Proceedings - International Conference on Pattern Recognition 28–31. <https://doi.org/10.1109/icpr.2004.1333992>.
10. Tian Y, Wang Y, Hu Z, Huang T (2013) Selective eigenbackground for background modeling and subtraction in crowded scenes. *IEEE Trans Circuits Syst Video Technol* 23(11):1849–1864. <https://doi.org/10.1109/TCSVT.2013.2248239>
11. Akilan T, Wu QMJ, Yang Y (2018) Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution. *Inf Sci (Ny)* 430–431:414–431. <https://doi.org/10.1016/j.ins.2017.11.062>
12. "Changedetection.net." <http://jacarini.dinf.usherbrooke.ca/dataset2012>
13. Barnich O, Van Droogenbroeck M (2009) "ViBE: A powerful random technique to estimate the background in video sequences," In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 945–948. <https://doi.org/10.1109/ICASSP.2009.4959741>.
14. L. Xu, F. Qi, R. Jiang, Y. Hao, and G. Wu, "Shadow Detection and Removal in Real Images: A Survey," *Citeseer*, 2006, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.1017&rep=rep1&type=pdf%5Cnpapers2://publication/uuid/F739EA7E-1488-4E22-88DC-DFD83F059410>
15. Agrawal S, Natu P (2020) Segmentation of moving objects using numerous background subtraction methods for surveillance applications. *Int J Innov Technol Explor Eng* 9(3):2553–2563. <https://doi.org/10.35940/ijitee.c8811.019320>
16. Khan SH, Bennamoun M, Sohel F, Togneri R (2016) Automatic shadow detection and removal from a single image. *IEEE Trans Pattern Anal Mach Intell* 38(3):431–446. <https://doi.org/10.1109/TPAMI.2015.2462355>
17. Abdusalomov A, Whangbo T (2017) An improvement for the foreground recognition method using shadow removal technique for indoor environments. *Int J Wavelets Multiresolution Inf Process*. <https://doi.org/10.1142/S0219691317500394>
18. Hofmann M, Tiefenbacher P, Rigoll G (2012) "Background segmentation with feedback: The Pixel-Based Adaptive Segmenter," In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 38–43. <https://doi.org/10.1109/CVPRW.2012.6238925>
19. Ebner M, Yuan X, Wang Z (2015) Single-image shadow detection and removal using local colour constancy computation. *IET Image Process* 9:118–126. <https://doi.org/10.1049/iet-ipr.2014.0242>
20. Sezgin M, Sankur B (2004) Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. *J Electron Imaging* 13:146–165. <https://doi.org/10.1117/1.1631315>

21. Pai CJ, Tyan HR, Liang YM, HY. M. Liao, and Chen SW (2003) “Pedestrian detection and tracking at crossroads,” In: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), 2, II–101. <https://doi.org/10.1109/ICIP.2003.1246626>.
22. Wren CR, Azarbayejani A, Darrell T, Pentland AP (1997) Pfnder: real-time tracking of the human body. *IEEE Trans Pattern Anal Mach Intell* 19(7):780–785. <https://doi.org/10.1109/34.598236>
23. Hassan MA, Malik AS, Nicolas W, Faye I (2015) Adaptive Foreground Extraction for Crowd Analytics Surveillance on Unconstrained Environments. In: Jawahar CV, Shan S (eds) *Computer Vision - ACCV 2014 Workshops*. Springer International Publishing, Cham, pp 390–400. https://doi.org/10.1007/978-3-319-16631-5_29
24. Stauffer C, Grimson WEL (1999) “Adaptive background mixture models for real-time tracking,” Proceedings. 1999 IEEE computer society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/cvpr.1999.784637>.
25. Haines T, Xiang T (2014) Background subtraction with dirichlet process mixture models. *Pattern Anal Mach Intell IEEE Trans* 36:670–683. <https://doi.org/10.1109/TPAMI.2013.239>
26. Karpagavalli P, Ramprasad V (2017) An adaptive hybrid GMM for multiple human detection in crowd scenario. *Multimed. Tools Appl.* 76(12):14129–14149. <https://doi.org/10.1007/s11042-016-3777-4>
27. Sengar SS, Mukhopadhyay S (2019) Moving object detection using statistical background subtraction in wavelet compressed domain. *Multimed Tools Appl* 79(9–10):5919–5940. <https://doi.org/10.1007/s11042-019-08506-z>
28. Liu Z, Huang K, Tan T (2012) Foreground object detection using top-down information based on em framework. *IEEE Trans Image Process* 21(9):4204–4217. <https://doi.org/10.1109/TIP.2012.2200492>
29. Jeevith SH, Lakshmikanth S (2021) Detection and tracking of moving object using modified background subtraction and Kalman filter. *Int J Electr Comput Eng* 11(1):217–223. <https://doi.org/10.11591/ijece.v11i1.pp217-223>
30. Zhang J, Guo X, Zhang C, Liu P (2021) A vehicle detection and shadow elimination method based on greyscale information, edge information, and prior knowledge. *Comput Electr Eng* 94:107366. <https://doi.org/10.1016/j.compeleceng.2021.107366>
31. Kim K, Chahidabhongse TH, Harwood D, Davis L (2005) Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*. <https://doi.org/10.1016/j.rti.2004.12.004>
32. Liu R, Ruichek Y, El-Bagdouri M (2019) Extended Codebook with Multispectral Sequences for Background Subtraction. *Sensors* 19(3):703. <https://doi.org/10.3390/s19030703>
33. Liao J, Wang H, Yan Y, Zheng J (2018) A Novel Background Subtraction Method Based on ViBe. In: Zeng B, Huang Q, El Saddik A, Li H, Jiang S, Fan X (eds) *Advances in Multimedia Information Processing – PCM 2017*. Springer International Publishing, Cham, pp 428–437. https://doi.org/10.1007/978-3-319-77383-4_42
34. Li W, Zhang J, Wang Y (2019) WePBAS: A Weighted Pixel-Based Adaptive Segmenter for Change Detection. *Sensors (Basel)* 19(12):2672. <https://doi.org/10.3390/s19122672>
35. Li S, Liu P, Han G (2017) Moving object detection based on codebook algorithm and three-frame difference. *Int J Signal Process Image Process Pattern Recognit* 10(3):23–32
36. Guo J, Wang J, Bai R, Zhang Y, Li Y (2017) A new moving object detection method based on frame-difference and background subtraction. *IOP Conf Ser Mater Sci Eng* 242:12115. <https://doi.org/10.1088/1757-899X/242/1/012115>
37. Chen J, Lu X, Ye M, Ming Z, Zhou F, Luo Y (2018) “A Moving Object Extraction Algorithm Based on Hybrid Background Subtraction and Pixel Mean Technique BT: Proceedings of the 2018 3rd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2018),” pp. 360–368. <https://doi.org/10.2991/amcce-18.2018.62>.
38. Dhingra G, Kumar V, Joshi HD (2021) Clustering-based shadow detection from images with texture and color analysis. *Multimed Tools Appl* 80(25):33763–33778. <https://doi.org/10.1007/s11042-021-11427-5>
39. Cucchiara R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.1109/TPAMI.2003.1233909>
40. Izadi M, Saeedi P (2008) “Robust region-based background subtraction and shadow removing using color and gradient information. <https://doi.org/10.1109/icpr.2008.4761133>.

41. Zhang W, Fang XZ, Xu Y (2006) "Detection of moving cast shadows using image orthogonal transform," In 18th International Conference on Pattern Recognition (ICPR'06) 1, 626–629. <https://doi.org/10.1109/ICPR.2006.441>.
42. Wu M, Chen R, Tong Y (2020) Shadow elimination algorithm using color and texture features. *Comput Intell Neurosci* 2020:2075781. <https://doi.org/10.1155/2020/2075781>
43. Murali S (2013) Shadow Detection and Removal from a Single Image Using LAB Color Space. *Cybern Inf Technol*. <https://doi.org/10.2478/cait-2013-0009>
44. Wang Y, Luo Z, Jodoin P-M (2017) Interactive deep learning method for segmenting moving objects. *Pattern Recognit Lett* 96:66–75. <https://doi.org/10.1016/j.patrec.2016.09.014>
45. Babae M, Dinh DT, Rigoll G (2018) A deep convolutional neural network for video sequence background subtraction. *Pattern Recognit* 76:635–649
46. K. Saarinen (1994) Image processing, analysis and machine vision, 35(1).[https://doi.org/10.1016/0165-1684\(94\)90202-x](https://doi.org/10.1016/0165-1684(94)90202-x).
47. Varghese A, Sreelekha G (2017) Sample-based integrated background subtraction and shadow detection. *IPSN Transact Comput Vision Appl*. <https://doi.org/10.1186/s41074-017-0036-1>
48. "SBU." https://www3.cs.stonybrook.edu/~cvi/projects/shadow_noisy_label/index.html
49. Montero VJ, Jung WY, Jeong YJ (2021) Fast background subtraction with adaptive block learning using expectation value suitable for real-time moving object detection. *J Real-Time Image Process* 18(3):967–981. <https://doi.org/10.1007/s11554-020-01058-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.