IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 28, 2016, accepted August 30, 2016, date of publication September 13, 2016, date of current version October 15, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2608847

An Evaluation of Background Subtraction for Object Detection Vis-a-Vis Mitigating Challenging Scenarios

SUMAN KUMAR CHOUDHURY, PANKAJ KUMAR SA, (Member, IEEE), SAMBIT BAKSHI, (Member, IEEE), AND BANSHIDHAR MAJHI, (Member, IEEE)
Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela 769008, India

Corresponding author: P. K. Sa (PankajKSa@nitrkl.ac.in)

This work was supported by the Science and Engineering Research Board, Department of Science & Technology, Government of India, under Grant SB/FTP/ETA-0059/2014.

ABSTRACT Background subtraction is a popular technique for detecting objects moving across a fixed camera view. The performance of this paradigm is influenced by various challenges, such as object relocation, illumination change, cast shadows, waving background, camera shake, bootstrapping, camouflage, and so on. In this paper, we present a synopsis on the evolution of the background subtraction techniques over the last two decades. The different ways of mathematical modeling are taken into consideration to categorize the methods. We also evaluate the performance of some of the state-of-the-art techniques vis-a-vis the challenges associated. Eleven different algorithms of background subtraction have been simulated on thirty-four image sequences collected from five benchmark datasets. For each image sequence, seven performance metrics are evaluated and an exhaustive comparative analysis has been made to derive inferences. The potential findings in the result analysis are presented for future exploration. The obtained image and video results are uploaded at https://sites.google.com/site/soaBSevaluation.

INDEX TERMS Video surveillance, object detection, background subtraction, background modeling, foreground extraction, background maintenance, shadow removal.

I. INTRODUCTION

Computational inefficiency, for the last few decades, has been a major bottleneck in processing videos in reasonable time. The recent advancement in parallel architectures has made feasible live-analysis of video data. It has also stimulated the researchers to develop more sophisticated and robust models that can deliver output under challenging conditions.

The objective of video surveillance is to extract essential information from a set of image sequences by automatically detecting and tracking the objects of interest followed by recognizing the relevant activities. Video Surveillance has enormous applications both in public and private sectors such as theft avoidance, crime hindrance, site visitors monitoring, combating in opposition to act of terrorism, land security, accident prediction, and many others. A generic surveillance framework comprises a set of cameras placed at strategic locations that are connected to digital computers to analyze the ongoing activities. This article concentrates on the very first phase of an automated surveillance framework: separating moving objects through background subtraction. Figure 1 illustrates few sample objects extracted as foregrounds from their respective frames and background model.

Background subtraction has been the most widely used approach to detect moving objects over the last two decades. In general, this framework is a three-stage process as shown in Figure 2. In the first stage, *i.e.* background initialization, either the first frame or few initial frames are taken into consideration to estimate a model of the background. Each successive frame is then compared with the established background to extract the moving objects during foreground extraction. The final stage, background maintenance, keeps updating the model to adapt any changes that may occur in the observed scene over the time. However, this framework is susceptible to various challenges such as cast shadow illumination, bootstrapping movement during initialization, background oscillation with varying periodicity across the view, the chromatic similarity between a foreground to its underlying background, gradual change in sunlight illumination over

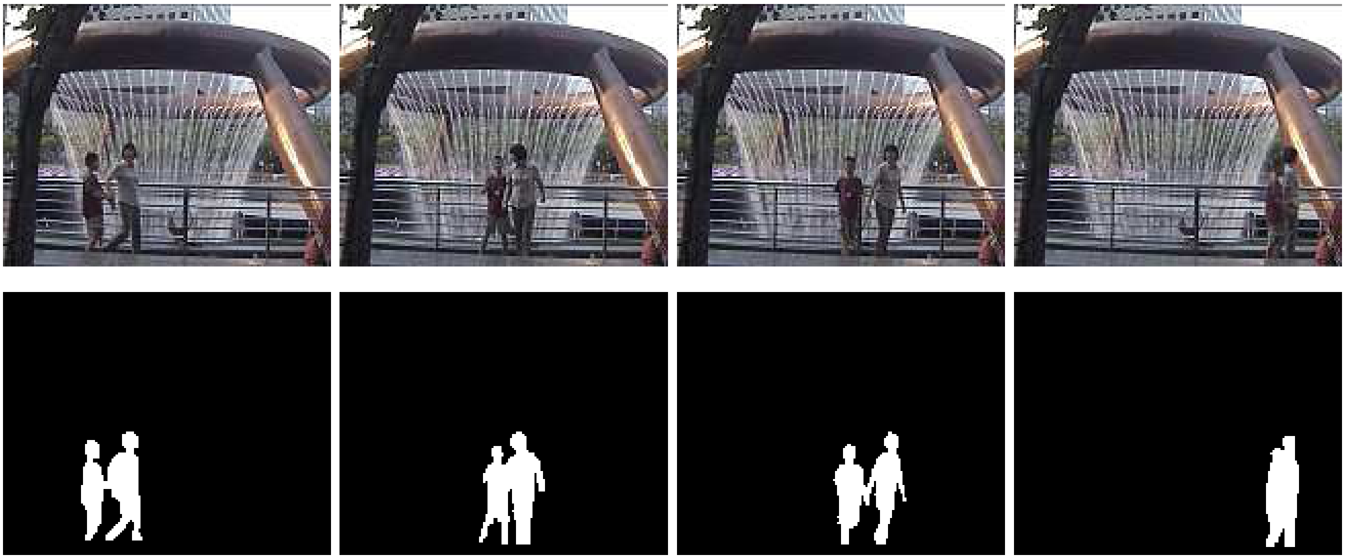

FIGURE 1. First row: four frames of the Fountain video; Second Row: Corresponding output, white pixels indicate foreground object.

the time, rapid varying illumination with cloud movement and switching the lights on/off. The subtle elements on each of these difficulties with their varying solutions are expounded in Section III.

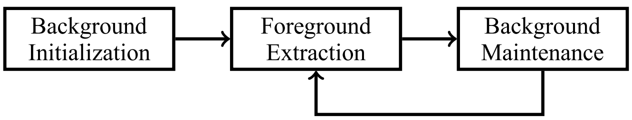

FIGURE 2. Primitive steps of background subtraction.

Quite a significant amount of work on background subtraction are available in literature [1]–[4]. In this paper, we provide an overview of the existing methods along with their solution strategy towards mitigating the challenges. We simulate eleven state-of-the-art methods on various image sequences collected from five benchmark datasets. An in-depth analysis of the results reveal some key findings in background subtraction methodologies. All results of this study are available at https://sites.google.com/site/soaBSevaluation.

The rest of this paper is organized as follows. Section II outlines the evolution of various mathematical models. Section III details the challenges as well as their solution strategies proposed over the years. Simulation statistics alongside the selected datasets, state-of-the-art methods, and performance measures are enumerated as well as the obtained results are analyzed in Sections IV and V. Finally, Section VI presents the concluding remarks.

II. THE GENESIS OF BACKGROUND SUBTRACTION

Effective background modeling and its periodic update are very much essential for accurate object detection. There exists plenty of literature in this field over the last two decades [5], [6]. However, there is no unique way to categorize these methods. In this section, we enumerate the evolution of various mathematical models of background subtraction to detect moving objects.

A. BASIC MODEL

The simplest way is to set the first frame as the background and subtract all subsequent frames to extract the foregrounds. However, an oscillating background cannot be adapted using a single frame. Furthermore, the very first frame of the sequence may contain moving objects that may falsely appear as background. The *frame difference* method uses the previous frame rather than the initial frame for subtraction purpose. This method well adapts the slow varying illumination; however, fails to update the background, when a moving object ceases its motion abruptly. Lai and Yung modeled the background using the arithmetic mean of pixel values over few temporal sequence [7]. The W4 system considers three tuples for subtraction over the initialization sequence; the minimum gray value, the maximum gray value, and the maximum intensity difference between two adjacent frames [8]. However, the inherent noise during image acquisition may substantially alter the intensity gap that leads to false positives and false negatives. All these methods are unimodal and therefore, an associate oscillatory background cannot be tailored employing a single-valued model.

B. STATISTICAL MODELS: SINGLE GAUSSIAN, MIXTURE OF GAUSSIANS, CODEBOOK

The initialization pixel sequence along the temporal axis is modeled using a univariate Gaussian distribution [9]. The multivariate distribution, for RGB color channels, is modeled as a product of three independent univariate Gaussian distribution, where each distribution is parametrized by the sample mean μ and standard deviation σ. The standard Z-score labeling is applied to extract the foreground pixels against the motion parameters. However, the single Gaussian model is unimodal and hence fails to accommodate the oscillating background. As an alternative, Stauffer and Grimson incorporate the mixture of Gaussians (MoG) to create a multi-modal background. Each pixel location is classified into $J \geq 1$ classes, where the unknown parameter J is chosen arbitrarily. The learning rate parameter is introduced to cope with varying illumination. The choice of learning rate parameter plays a major role; low learning rate fails to tackle sudden illumination variations, whereas higher rate includes the slow-moving objects in the background [10], [11]. Zivkovic and Hayden, in their work, proposed a solution to choose the correct number of background classes at each location based on their sample variation over the frames [12], [13]. In another work, Kaewtrakulpong and Bowden changed the background update equations in the original MoG model to address the rapid illumination variations [14]. The codebook scheme uses few statistical attributes to encode each background location such as the minimum and maximum intensity for a pixel over the frames, frequency-of-occurrence for each codeword, frame number at which the codeword has first time occurred, frame number at which the codeword last appears, and the maximum frame gap during which the codeword remains missing [15], [16]. Fernandez-Sanchez *et. al.* follow the same principle to initialize the background; moreover, they incorporate the depth cue as an additional model parameter to strengthen the discriminating ability of the developed model [17].

C. NON-PARAMETRIC MODEL

The temporal pixel sequence at any location might not follow the default Gaussian distribution. The kernel density estimation (KDE) techniques are applied in those scenarios, where the underlying distribution is unknown. In particular, these algorithms take ample training samples to converge to the underlying density function [18], [19]. The KDE based methods strongly depend on the suitable choice of kernel bandwidth that must have finite local support. Moreover, the bandwidth is inversely related to the number of frames adapted for background initialization. A narrow bandwidth results in a jagged density estimation, whereas an extensive one leads to over-smoothed distribution [20]. Piccardi and Zen, in their work, apply the median of the absolute difference between adjacent frames to estimate the kernel bandwidth [21]. The varying waving periodicity in the case of oscillating background appeals to approximate kernel bandwidth for each model location across each of R, G, B color channel. In another work, the mean-shift paradigm is chosen to estimate the kernel bandwidth with fewer training samples [22]. Elgammal *et al.* apply a fast Gauss transform to reduce the overall response time of density computation [23]. Mittal and Paragios, in their work, model the dynamic background using optical flow, and the feature uncertainty is resolved using a KDE technique [24]. Parag *et al.* suggested a boosting based ensemble learning to select appropriate features for the KDE based methods [25].

D. NON-RECURSIVE BUFFER BASED SUBTRACTION

Lo and Velastin store the recent pixel history in a finite buffer to represent the model location [26]. The significant difference between the current pixel and buffer median decides if it were a foreground; else, the new background is enqueued inside the buffer. The first-in-first-out strategy is applied to tackle the situation when the overflow condition is encountered. Subsequently, Cucchiara *et al.* prefer the medoid rather than the median statistics to take the appropriate decision [27], [28]. In another work, the background is modeled using a linear predictive model through Wiener filtering [29]; the covariance of pixel sequence estimates the filter coefficients. This work is further extended in a relevant subspace via PCA [30], [31]. Wang and Suter use the notion of consensus to model the background. Additionally, two algorithms are suggested to deal with rapid varying illumination and background relocation [32], [33].

E. FUZZY MODEL

Fuzzy principle can be incorporated to address the decision uncertainty during foreground extraction [34]. Zhang and Xu incorporate the fuzzy Sugeno integral to model the underlying scene under observation [35]. Subsequently, the Choquet integral is preferred over the former one that yields comparatively better accuracy [36]. Azab *et al*. fuse the edge information along with the color and texture features to model the background, where the Choquet integral is applied to extract the foreground pixels. Bouwmans *et al.* proposed another type-2 fuzzy technique to take the classification decision [37], [38]. Kim and Kim apply the fuzzy color histogram to model the waving background [39].

F. LEARNING MODEL

A classifier (for example, neural network) is used to train the underlying density distribution of the pixel sequence and decide the nature of the next picture element. Culibrk *et al.* train a probabilistic, multi-layered, feed-forward neural network with 124 neurons to create a background model. A Bayesian classifier is then employed to separate the non-stationary pixels [40]. Maddalena and Petrosino design a self-organization map network in which each background location is represented by a set of learned weight vectors [41], [42]. Moreover, a spatial coherence paradigm is introduced to reduce false alarms.

G. LOW-RANK SPARSE DECOMPOSITION

Subspace learning models are also introduced in the field of background subtraction. The Robust Principal Component Analysis (RPCA) is applied to decompose the video frames into a low-rank background matrix and a sparse foreground matrix [43], [44]. Wright *et. al.*, in their work, incorporated a L_1-norm on the sparse matrix such that the background frames are linearly correlated to each other [45]. In another work, the Total Variation regularization constraint is incorporated to handle the noisy data [46]. The low-rank matrix (background model) assumes to capture any variation that has been observed at the underlying scene over the time. However, The RPCA paradigm considers the entire image sequence as a vectorized data matrix and therefore has high memory overhead [47]. Sobral *et. al.* suggested an incremental tensor subspace learning that builds the low-rank matrix using few initial temporal sequence only, and periodically update the model with subsequent frames [48]. In another work, a Sparse Outlier Iterative Removal (SOIR) algorithm is employed to model the background scene, where a cyclic iteration process is suggested to separate foreground pixels [49], [50].

H. SHADOW REMOVAL MODEL

Shadow darkens the scene illumination, and hence, the underlying region falsely appears as foreground. The literature includes two different ways to tackle this situation. The former group suggests various invariant color models [27], [51] to nullify the shadow effects, whereas another set of algorithms prefer the texture feature that remains indifferent in the presence of shadow [52], [53]. Wang and Suter, in their work, apply the normalized *RGB* color space to model the background; however, it has been observed that the normalized *RGB* is very much noisy in case of low intensity [54]. Cucchiara *et al.* apply the *HSI* model to suppress the shadow illumination and use a median filter to selectively update the established model [27]; a second validation is further applied using both invariant color and texture pattern of the underlying scene [28]. Huang *et al.* incorporate the color and color co-occurrence features to model the static and waving background respectively [55]. Huerta *et al.* suggest a two-stage approach to counter the shadow illumination [56]. The first stage combines the gradient details and color information to detect the probable shadow pixels. A second validation is further applied based on the temporal and spatial analysis of chrominance measure, brightness content, texture distortion, and diffused bluish effect. Zhou *et al.* consider multiple cues such as motion details, object location, its shape, and color feature to detect the objects in motion [57].

I. POST-PROCESSING REFINEMENT

Foreground extraction may be erroneous owing to the cluttered background and the inherent sensor noise during image acquisition. It may so happen that a few portion of foreground pixels may be wrongly identified as the background and vice-versa. A post improvisation module should be incorporated to minimize such false alarms [58]. The median filtering is a suitable tool to reduce such false positives. Again, some methods apply connected component analysis to attach the disjointed regions. The size constraint as per the objects of interest can be incorporated to eliminate small foreground pixels. Many authors prefer morphological post-filtering for such improvisation. Morphological *Opening* is applied to reduce the scattered noise pixels. The *closing* operation connects the disjointed pixels. Moreover, the morphological *filling* can be applied to fill the camouflage gap.

The literature includes a number of articles on the use of background subtraction in identifying the moving entities. *Parametric* models are based on their underlying assumptions; the appropriate parameter selection can be cumbersome and moreover, it may vary with different scene structures. On the other hand, *non-parametric* models are more reliable, however, requires a long pixel history to estimate the underlying density function. *Pixel-based* methods usually apply the color feature to compare the pixel intensities at the same location over the frame sequence, whereas *block-based* methods consider the inter-pixel neighborhood characteristics, partition the image into several blocks, and apply both color and texture cues to decide the pixel behavior. The *unimodal* background outputs significant false positives in case of uninteresting waving motion that can be tackled by the *multi-modal* background at the price of higher space complexity. The *recursive* models update the model parameters in an iterative fashion, and thereby fast enough to deploy in real time applications. On the contrary, the *non-recursive* techniques store the recent pixel information inside a buffer to model the background. The latter one well adapts the gradual illumination variations at the cost of high memory overhead.

III. CHALLENGES AND MITIGATIONS

Moving object detection through background subtraction is usually challenged by a number of factors. The strength and weakness of any background subtraction algorithm are assessed by observing the efficiency with which it counters the challenges. In this section, we detail the possible challenges along with their solutions proposed over the years.

A. GRADUAL ILLUMINATION CHANGE

Visibility of outdoor scenes is, in general, affected by the problem of slow illumination change. The sunlight illumination varies gradually over the day and thereby inducing a deviation in the modeled background. As a result, faulty foreground pixels appear, even though, no real object movement has occurred.

Earlier schemes apply the recursive paradigm to model the underlying background. Motion parameters, such as *mean* and *variance*, are recursively updated with each forthcoming background pixel. In this principle, only two parameters per location are required to model the entire scene. Such recursive update over a longer period may include the distant past pixel contribution. The so formed model parameters may get more skewed towards the old values. It may so happen that a true background pixel may significantly differ against such biased distribution. Such misclassification adversely impacts the ensuing decisions for a more drawn out period.

The non-recursive methods, on the other hand, store only the recent pixel history in a finite buffer to model each location. The subsequent pixels are then compared with the buffer statistics such as median, medoid *etc*, to take necessary decision. A forthcoming background pixel is included in the buffer as long as the overflow condition is not encountered, or replaced by an existing buffer value when its size is full.

B. UNINTERESTING BACKGROUND OSCILLATION AND CAMERA SHAKE

The waving of leaves, vacillating of banners, the fluttering of flags, water stream in wellsprings are few real life instances which are usually independent of motions under consideration. In addition, the camera may also undergo motion due to external forces.

Earlier background models are inherently unimodal as each background location is defined either by a single value or by a finite range. As an alternative, multi-label background models are in use for the last one and half decade. In such modeling, multiple classes are assigned for the background coordinates so that the waving patterns can be sampled without any loss.

The earlier multi-modal systems assign an equal number of classes for each pixel location. However, the waving pattern may not be uniform across the background. As a result, the equal class distribution, without any knowledge of scene structures, may fail to accommodate all possible variations. The recent state-of-the-art methods first learn the oscillation periodicity for each pixel location with sufficient training samples, and then, assign the required number of classes across the model.

C. SHADOW AND REFLECTION

Shadow darkens the underlying region, whereas diffuse reflection brightens the scene visibility. In other words, a shadow can be interpreted as a scaled down value of luminance, whereas diffuse illumination as a scaled up value. In either case, the resulting deviation yields unnecessary false positives. These problems are usually suffered by methods that are solely based on luminance features. In case of *RGB* color space, each of the channels is a linear combination of both luminance and chrominance component. The shadow removal strategy demands an invariant pixel representation that should be independent of luminance channel. In other words, the so formed pixel data structure should be a function of the chrominance measure only. There exist two different ways of addressing the shadow illumination issues. Chromatic channels such as *Hue* from *HSV*, *a*, *b* from *Lab*, *Cb*, *Cr* from *YCbCr* color channel, and so forth, are applied to nullify the illumination factor. Another set of methods applies a battery of texture features that remain invariant to illumination.

D. BOOTSTRAPPING

In the best case, there would be no object movement at the observed scene during background modeling. Foreground movements during initialization create faulty classes in the developed model, known as the bootstrapping problem. Once, the foreground objects leave their location, the rearward background appears as non-stationary in rest of the frames. On the contrary, the same foreground object, or look-alike objects, when pass across the underlying region, may remain undetected due to the match with the faulty background class.

Usually, the appearance frequency of a background class is taken into consideration to figure out the bootstrapping problem. A moving object cannot remain stationary at the same location for a longer period. In other words, the appearance frequency of such faulty classes is very low as compared to that of a true background class. Once the background modeling is over, a suitable outlier labeling method can be applied to remove the low-frequency classes.

E. BACKGROUND DISPLACEMENT

A background model might change after initial training owing to various scenarios such as parking a vehicle, replacing an old refrigerator with a new one with varying shape and size, shifting chairs to another room *etc*. The above scenarios can be generalized into three different categories, as given below.

Case 1: When a new background object is introduced into the scene, it is quite apparent that the number of objects in the foreground increases by one.

Case 2: When an existing background object is removed, it creates a void space whose pixel distribution is not at par with the prevailing background, and thus introducing a new foreground object that actually does not exist.

Case 3: The relocation of existing background objects could be interpreted as a combination of the above two cases. An existing background object is removed from its current position (Case 2) and placed at another location (Case 1), and creating an illusion of two new objects, which is in actual one object.

The developed system should be robust enough to update such changed location as background. Usually, the visibility duration of each object is taken into consideration to alleviate this issue. In particular, a foreground object that remains stationary at a location for a longer time period will be incorporated into the background model. On the contrary, an existing background class that remains absent for sufficient frame period will be removed from the model.

F. CAMOUFLAGE

Foregrounds significantly deviate from the modeled background in terms of their visual appearance. Most of the existing schemes incorporate the intensity and color difference as a measure to identify the non-stationary pixels. However, color alone cannot solve the detection task. It may so happen that the color of a moving object may match to its rearward background. As a result, the pixel difference lies within the set threshold that yields false negatives. Recent schemes incorporate the texture or gradient information along with the color feature to strengthen its ability to detect the foreground objects more accurately. The only case when it may fail lies with same color and texture values at both foreground and background points; however, such scenarios are very rare.

G. SUDDEN ILLUMINATION VARIATION

Usually, the number of background pixels at any instant of time exceeds the number of foreground pixels. However, the rapid illumination variation alters the entire visibility of the scene, and maximum background pixels significantly deviate from their original value. The indoor illumination is usually disturbed by the lights on/off at times, whereas the outdoor environment fluctuates in the presence of cloud.

The simplest solution is to reinitialize the model as soon as the fast variation is observed. The usual background update strategy discussed in Section III-E can also be applied at the expense of few successive frames so that the model can be relabeled automatically. Detection result during this frame interval may be faulty; however, the outcome is more reliable at the expense of this small frame gap. Methods based on Gaussian mixture model update the background using a learning rate parameter. Another set of algorithms models the variation transformation as regression polynomial of probable background coordinates.

IV. SIMULATION SETUP

In our work, we simulate some state-of-the-art algorithms on several image sequences. The outcomes consequently acquired, are investigated and the findings are summarized in the next section. In this section, we detail the standard datasets and various performance measures used for the evaluation.

A. DATASETS

Benchmark datasets alongside their ground-truth annotations are crucial for both qualitative and quantitative analysis of any algorithm. In this work, we simulate several image sequences collected from five benchmark datasets namely, Wallflower, I2R, Carnegie Mellon, Change detection (CDW), and Background Models Challenge (BMC). The selected image sequences alongside the underlying challenges are listed in Table 1.

Wallflower: We simulate five out of eight image sequences from Wallflower dataset [29]. Besides, one ground-truth image is provided with each sequence, which is compared on a pixel-by-pixel basis.

Camouflage: In this video, a man strolls over the screen of a computer. The color of his shirt matches to the moving interlacing bars on the computer monitor. Towards the end, the person casts shadow on the side wall.

Bootstrapping: This video is recorded inside a cafeteria. Foreground movements can be observed from the very first frame of the sequence.

TimeOfDay: The gradual variation of daylight illumination over a day is portrayed in the *TimeOfDay* sequence. The video demonstrates a moderately lit vacant room being brightened gradually and uncovering various items present in it. Towards the end, a man enters the room and sits on a couch.

LightSwitch: The lights are initially off with no moving objects inside the room. After a while, a person comes, switches the light on, and leaves the room.

TABLE 1. Selected videos and the underlying challenges.

Challenge	Dataset	Video
Background relocation	CDW/ Intermittant Object Motion	Parking
		Sofa
		WinterDriveAway
Bootstrapping	Wallflower	Bootstrap
	BMC/ Real Video	WonderingStudents
CameraShake	CDW/ Camera Jitter	Boulevard
Camouflage	Wallflower	Camouflage
Gradual illumination variation	Wallflower	TimeOfDay
Shadow	CDW/Shadow	Backdoor
		Bunglaows
		BusStation
Sudden illumination variation	I2R	Lobby
	Wallflower	LightSwitch
Background oscillation	CDW/ Dynamic background	Canoe
		Fountain01
		Fountain02
		Overpass
	I2R	Campus
		Fountain
		WaterSurface
	Wallflower	WavingTrees
Camera Shake, Shadow	Carnegie Mellon	Carnegie Mellon
Bootstrapping, Shadow	I2R	ShoppingMall
	BMC/ Real Video	TrainInTunnel
Camera Shake, Shadow, Bootstrapping	CDW/ Camera Jitter	SideWalk
		Traffic
Background Oscillation, Camouflage	I2R	Curtain
Background Oscillation, Illumination change, Bootstrapping	I2R	Escalator
Bootstrapping, Shadow, Camouflage	I2R	Hall
	BMC/ Real Video	BigTrucks
Sudden illumination variation, Background Oscillation, Camera shake, Bootstrapping	BMC/ Real Video	TrafficDuringWindyDay
Sudden illumination variation, Background noise, Shadow, Background oscillation	BMC/ Real Video	RabbitInNight
Background oscillation, Background relocation, Bootstrapping, Shadow, Camouflage	BMC/Real Video	BewareOfTrains

WavingTrees: It demonstrates an oscillating background that includes a person walking across a swaying tree.

I2R: We simulate seven out of nine image sequences from I2R dataset [55]. Each sequence is again associated with 20 hand segmented images. Accordingly, the result is computed across these 20 ground-truth annotations unlike single ground-truth in Wallflower dataset.

Lobby: This video is captured inside a room with five light sources illuminating the scene. It can be observed that switching different lights on/off at various times alters the visibility of the prevailing background.

Campus: This sequence delineates an open air scene where the waving trees yield uninteresting background movement.

Fountain: Pedestrians are walking in front of a water fountain. Any detection algorithm should incorporate the water flow in the background model, otherwise, it results in false positives.

WaterSurface: This is another instance of uninteresting background movement, where the sea waves, if not dealt properly, will give rise to significant false positives.

Curtain: It portrays an instance of both camouflage and waving background. The fluttering curtain should be absorbed in the model. Furthermore, the disguise issue can be observed owing to the attire similarity between the foreground to that of underlying background.

Escalator: This video is recorded in a subway station. The pedestrians, as well as the escalators, are in motion from the very first frame itself. It illustrates the issue of bootstrapping as well as background oscillation. In addition, the variation in light illumination can be observed over the time.

Hall: It delineates pedestrian motion from the very first frame alongside their cast shadow. Furthermore, the camouflage issue comes into picture owing to the similarity in pedestrian clothing with various parts of the background.

Carnegie Mellon: This dataset [59] has only one image sequence that contains 500 TIF frames along with their hand segmented annotations. The scene is recorded with nominal camera movement, which is of 18 pixels on average.

Background Models Challenge (BMC): This dataset is partitioned into three different categories namely, *Learning mode, Synthetic videos* and *Real videos*. In this work, we simulate six out of eight sequences from the *Real videos*.

WonderingStudents: The bootstrapping movement is well reflected from the very first frame of the sequence.

BewareOfTrains: This sequence portrays a number of challenges over the frames. The swaying leaves yield background oscillation, whereas the chromatic match between foreground cars and rearward background raises the camouflage issue. Furthermore, the underlying shadow of the moving vehicle alters the scene appearance, and the trains journey, running $\rightarrow$ steady $\rightarrow$ running, appeals to update the background model.

BigTrucks: The truck is big in size, homogeneous in color, and very close to the camera. The interior pixels, while the truck is in motion, may appear as stationary due to pixel homogeneity across the neighborhood. Movement during initialization, cast shadow, and the chromatic match between the truck and side wall are few other issues.

RabbitInNight: The quantization noise in this low-resolution video appears as a continuous motion over the temporal sequence. In addition, the rapid change in light illumination and the shadow of the walking pedestrian are other concerns need to be taken care.

TrafficDuringWindyDay: The camera shake and waving trees are the reasons for uninteresting background oscillation, whereas the car movement during initialization raises the bootstrapping issue. Furthermore, the cloud movement changes the appearance of the prevailing view.

TrainInTunnel: The bootstrapping motion of an individual and his shadow can be well visualized in this sequence.

Change Detection.Net (CDW): This dataset is partitioned into eleven categories with several image sequences in each category [60], [61]. Furthermore, each video is associated with their hand segmented annotations. In some sequences, the ground-truth contains binary results from part of an image only rather than the entire frame; in those videos, the results are generated across the interest region only. In particular, we simulate thirteen videos from CDW dataset.

Parking, WinterDriveAway: In both sequences, the cars are initially parked for some time and therefore behave as stationary against the developed model. After a while, one person drives a car away from the scene and thereby creating a ghost space that may falsely appear as foreground. Furthermore, the cloud movement in the *Parking* sequence alters the scene visibility in terms of rapid varying illumination.

Sofa: The primary focus in this video is a sofa and the various items placed on or near by it. Over the time, few background objects are either shifted to another location or taken away from the camera view. A background model should take care of all such movements.

Backdoor, BusStation, Bungalows: Pedestrian motion in the first and second sequences, and moving vehicle in the third sequence cast shadow that darkens the true intensity. Alongside, the waving tree in the first video demands a multimodal background.

Canoe, Fountain01, Fountain02, Overpass: The river flow in the first video, the fountain water in the second and third videos, and the waving trees in the fourth video need to be absorbed with the background.

Sidewalk, Traffic: The camera oscillation in both sequences result in a dynamic background that is actually independent of motion under consideration.

B. STATE-OF-THE-ART COMPARISON

In our work, we simulate eleven state-of-the-art methods for comparison analysis —(i) Block-based classifier cascade with probabilistic decision integration (`BCCPDI`, [62]), (ii) Pfinder: real-time tracking of the human body (`Pfinder`, [9]), (iii) Fuzzy integral for moving object detection (`FuzzyIntegral`, [63]), (iv) Self organizing background subtraction (`SOBS`, [41]), (v) Improved adaptive background mixture model (`IABMM`, [14]), (vi) Multilayer background subtraction based on color and texture (`MultiLayer`, [64]), (vii) Pixel-based adaptive segmenter (`PBAS`, [65]), (viii) Fast principal component pursuit via alternating minimization (`FPCP`, [43]), (ix) GoDec: Randomized low-rank & sparse matrix decomposition (`GoDec`, [44]), and two variants of `ViBe` [66]: (x) `ViBeRGB`, based on RGB color space and (xi) `ViBeGray`, based on gray color space.

C. PERFORMANCE MEASURES

Background subtraction can be interpreted as a binarization process in which each pixel of the current frame can either be labeled as background (black) or foreground (white). The efficacy of such algorithms are evaluated by computing the number of correctly identified motion pixels (true positives TP), the number of correctly labeled background pixels (true negatives TN), the number of pixels that are incorrectly detected as foreground (false positives FP) or wrongly labeled as background (false negatives FN). A confusion matrix is created using these four parameters as shown in Table 2.

The following seven evaluation metrics are computed using the above four parameters.

TABLE 2. Confusion matrix for background subtraction.

		Actual value (Ground truth fact)	
Predicted value (Predicted by simulation)		Foreground	Background
	Foreground	TP (True Positive)	FP (False Positive)
	Background	FN (False Negative)	TN (True Negative)

PCC, percentage of correct classification, is the proportion of correctly detected pixels over total image pixels under consideration.

$$PCC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

Specificity, known as true negative rate, measures the percentage of correctly identified negative samples over the actual number of negatives present in the ground-truth.

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (2)$$

False Positive Rate (FPR), known as fall-out, outputs the proportion of false positives that are retrieved by any algorithm, out of the total number of negative samples in the ground-truth.

$$FPR = \frac{FP}{FP + TN} \times 100 \quad (3)$$

False Negative Rate (FNR), known as miss rate, outputs the proportion of false negatives that are retrieved by any algorithm over the total number of positive samples in the ground-truth.

$$FNR = \frac{FN}{FN + TP} \times 100 \quad (4)$$

Recall, known as detection rate, is the ratio of number of true positives to the total number of positive examples annotated in the ground-truth.

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (5)$$

Precision, known as positive prediction, is the ratio of number of true positives to the total number of foreground pixels detected by any algorithm.

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (6)$$

F_1 **measure**, known as figure of merit, is the harmonic mean of Precision and Recall. Higher is the score, better is the efficacy.

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

These seven measures lay the basis of our analysis that we discuss in the next section.

V. RESULTS AND DISCUSSIONS

We simulate eleven state-of-the-art algorithms (listed in Section IV-B) on 34 image sequences collected from five benchmark datasets (enumerated in Section IV-A). The tabular results of the above seven evaluation metrics are listed in Tables 3, 4, 5, 6, 7, 8, and 9. Furthermore, the obtained binary images and video results of the test sequences are uploaded at https://sites.google.com/site/soaBSevaluation. For the readers' perusal, the variation in results distribution over the simulated image sequences are compared in Figures 3, 4, 5, 6, 7, 8, and 9 for all state-of-the-art algorithms; the vertical red bar demonstrates the results variation across the simulated videos, whereas a green rectangle in each red bar depicts the average performance of the corresponding approach. The following paragraphs summarize the in-depth analysis of the obtained results with the perspective of various challenges and performance metrics.

A. ANALYSIS FROM CHALLENGE PERSPECTIVE

1) ANALYSIS ON BACKGROUND DISPLACEMENT

We simulate three image sequences that depict the background relocation scenario, namely *Parking, Sofa*, and *WinterDriveAway*. It has been observed that `SOBS`, `FPCP`, `GoDec`, and `BCCPDI` have good recall rate whereas `IABMM`, `Multilayer`, `Pfinder`, `VibeRGB`, `VibeGray` have good precision rate. Background displacement strongly depends on the scene under observation. It is quite impractical to set a predefined threshold of time beyond which all stationary foregrounds can be absorbed into the background. On the contrary, it is very tough to strict an absence duration threshold beyond which an existing background class will be removed from the developed model. These two parameters have to be varied with respect to the underlying environment. The scene knowledge along with the information of possible stationary objects have to be learned to reduce such false alarms.

2) ANALYSIS ON BOOTSTRAPPING

Both *Bootstrapping* and *WonderingStudents* videos well reflect the bootstrapping scenario. All methods except `IABMM` have satisfactory output. Bootstrapping can be considered as a special case of background relocation wherein the knowledge of possible objects, their size, average halt duration *etc*, have to be learned over the initialization sequence to remove the faulty background classes from the developed model.

3) ANALYSIS ON CAMERA SHAKE

Camera oscillation can be observed in *Boulevard, SideWalk, Traffic,* and *Carnegie Mellon* sequences. `IABMM` has very poor recall rate, whereas the `Pfinder`, `FPCP`, `GoDec` marginally drop the precision rate. The oscillation periodicity owing to camera-shake needs to be learned with sufficient initialization frames.

TABLE 3. Comparative analysis of percentage of correct classification (PCC).

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	82.07	89.28	87.44	81.44	86.47	86.69	**92.10**	75.16	73.40	88.60	89.01
Sofa	**98.18**	95.18	93.97	97.55	94.39	95.55	95.95	95.63	95.30	95.70	95.80
WinterDriveAway	94.12	97.72	98.74	90.77	98.61	98.74	**98.83**	88.00	58.28	96.83	96.63
Bootstrapping	88.84	91.31	90.87	87.83	87.79	**94.55**	87.78	91.63	91.54	87.45	88.55
WonderingStudents	99.21	99.21	**99.39**	98.82	98.86	99.02	99.21	98.35	97.96	99.36	99.32
Boulevard	**98.01**	89.79	95.24	85.47	94.71	97.38	97.69	90.95	94.31	95.37	95.16
Camouflage	85.33	94.69	96.70	87.38	78.46	**97.44**	89.67	50.97	50.66	89.78	90.04
TimeOfDay	75.05	95.16	86.91	91.73	94.43	**98.51**	94.81	96.02	77.04	93.30	93.97
Backdoor	**99.54**	98.19	98.47	94.52	98.02	99.40	98.09	92.78	91.84	98.48	98.60
Bunglaows	97.57	95.84	95.51	97.07	94.37	**97.72**	95.97	93.92	93.71	95.66	96.22
BusStation	**97.99**	95.58	95.69	95.63	94.29	95.95	96.73	96.10	95.85	96.29	96.65
Lobby	94.68	96.04	95.77	88.72	98.10	**98.43**	95.04	88.81	88.91	94.52	94.04
LightSwitch	77.80	17.80	22.22	28.55	**85.23**	77.36	16.82	62.93	23.00	36.65	32.47
Canoe	98.08	91.87	97.61	87.05	94.09	96.67	**98.68**	83.60	82.61	96.61	97.41
Fountain01	98.99	97.58	98.84	95.87	99.34	99.65	**99.87**	90.70	90.31	99.36	99.19
Fountain02	99.89	99.62	99.71	96.20	99.69	99.87	**99.91**	91.97	99.64	99.75	99.78
Overpass	99.10	97.68	97.35	94.56	98.15	98.79	**99.29**	93.52	98.13	98.93	98.97
Campus	98.93	92.66	97.71	86.70	97.67	**99.28**	98.69	92.20	91.18	97.67	97.69
Fountain	98.13	98.05	98.20	91.96	97.82	**98.75**	98.38	94.66	91.90	97.54	97.49
WaterSurface	**98.91**	98.03	98.09	97.37	93.33	94.42	97.13	92.99	89.71	98.06	98.05
WavingTrees	96.44	92.96	92.54	90.14	88.95	**97.33**	90.96	58.90	58.03	83.08	96.77
Carnegie Mellon	**99.69**	97.32	98.70	98.69	98.62	99.30	99.48	97.43	97.02	98.93	98.92
ShoppingMall	96.78	96.32	96.09	93.51	95.72	**97.25**	96.44	95.50	95.39	96.03	96.20
TrainInTunnel	94.51	96.34	95.26	93.66	95.31	95.25	95.96	94.88	94.79	**96.70**	96.64
SideWalk	95.83	85.70	94.12	91.69	93.95	95.29	**97.24**	75.80	75.60	95.09	94.97
Traffic	96.06	92.27	92.37	89.85	94.14	**96.24**	95.94	88.26	85.38	94.62	94.64
Curtain	**98.77**	95.83	95.41	97.61	92.69	94.11	95.71	91.18	80.67	96.31	96.59
Escalator	95.92	93.48	83.97	88.40	95.98	95.21	96.00	92.30	91.71	97.02	**97.12**
Hall	96.08	95.91	95.86	94.89	95.29	95.98	95.57	95.48	94.64	96.32	**96.43**
BigTrucks	93.08	93.24	92.68	79.12	92.77	93.47	**93.50**	92.61	92.52	90.35	89.53
TrafficDuringWindyDay	**98.10**	93.88	97.68	82.31	96.82	96.58	97.23	79.53	77.68	93.13	92.44
RabbitInNight	98.64	99.34	98.75	91.78	99.11	99.33	**99.56**	82.97	81.62	99.29	99.28
BewareOfTrains	92.83	93.71	92.75	88.71	94.44	**94.88**	94.62	92.04	91.57	94.09	93.82
Average	94.94	92.65	93.05	89.56	94.47	**96.19**	93.90	87.81	84.60	93.54	93.89

TABLE 4. Comparative analysis of specificity.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	93.18	98.97	99.67	83.69	99.98	**99.99**	99.53	72.69	70.49	99.56	99.49
Sofa	99.23	99.66	98.21	99.44	**99.98**	99.86	99.50	98.22	97.80	99.93	99.91
WinterDriveAway	94.62	98.31	99.55	91.00	**99.96**	99.87	99.29	88.08	58.41	97.50	97.27
Bootstraping	98.60	98.18	96.73	91.67	**99.78**	98.68	99.75	96.83	96.97	98.17	97.59
WonderingStudents	99.30	99.50	**99.92**	99.01	**99.92**	99.20	99.31	98.49	98.08	99.67	99.59
Boulevard	98.90	91.37	98.07	85.58	99.11	99.80	99.32	91.68	**100.00**	98.04	97.32
Camouflage	91.55	95.16	**98.15**	75.02	97.36	94.41	88.36	11.31	11.09	93.16	91.20
TimeOfDay	73.42	99.98	86.65	97.67	99.99	99.05	99.66	98.30	77.62	**100.00**	**100.00**
Backdoor	99.85	99.26	99.90	95.00	**100.00**	99.90	98.55	93.26	92.28	99.86	99.76
Bunglaows	97.63	97.82	97.98	98.93	**98.98**	98.25	96.41	95.34	95.09	97.82	97.82
BusStation	98.88	99.07	99.64	96.95	**99.97**	99.79	99.82	97.49	97.17	99.82	99.78
Lobby	94.96	97.44	97.46	89.06	**99.98**	99.85	96.55	89.57	90.55	96.25	95.63
LightSwitch	73.55	14.56	7.99	16.43	**98.83**	73.98	5.17	59.69	20.58	36.51	29.98
Canoe	99.25	93.05	99.83	87.06	98.63	99.92	**99.99**	83.48	82.41	99.37	99.12
Fountain01	99.03	97.66	98.95	95.92	99.46	99.71	**99.92**	90.73	90.33	99.47	99.29
Fountain02	99.95	99.78	99.96	96.26	99.95	99.97	99.99	91.97	**100.00**	99.97	99.94
Overpass	99.41	98.35	98.19	94.74	99.75	99.86	99.96	93.75	**100.00**	99.79	99.70
Campus	99.06	93.38	99.06	86.69	98.99	99.49	**99.64**	92.46	91.40	99.25	98.96
Fountain	98.42	99.20	99.75	92.39	**99.88**	99.61	99.56	94.80	91.92	99.48	99.27
WaterSurface	99.17	99.68	**99.97**	98.13	99.95	99.87	99.54	93.03	89.40	99.95	99.91
WavingTrees	96.05	91.23	**98.46**	86.07	98.22	96.15	98.00	50.01	48.62	97.97	96.96
Carnegie Mellon	99.78	97.87	99.77	98.80	**99.93**	99.84	99.85	97.56	97.14	99.56	99.38
ShoppingMall	97.29	99.35	99.44	94.65	**99.70**	98.81	98.70	96.79	96.64	99.47	99.26
TrainInTunnel	94.89	98.33	97.93	94.38	**99.20**	98.50	98.00	95.72	95.56	98.82	98.57
SideWalk	98.56	88.38	97.70	93.74	98.53	**99.00**	98.83	76.84	76.59	98.90	98.52
Traffic	97.53	93.88	94.93	90.47	**98.82**	97.46	96.95	89.16	86.17	97.23	96.89
Curtain	99.52	99.19	99.29	98.73	99.89	99.80	99.77	90.83	79.14	**99.92**	99.90
Escalator	96.53	94.55	84.62	88.73	**99.50**	95.96	99.30	92.76	92.08	99.45	99.22
Hall	96.52	98.58	99.61	95.60	**99.68**	99.33	98.92	96.55	95.54	99.53	99.39
BigTrucks	97.81	96.35	97.63	80.24	**99.61**	98.81	96.47	93.82	93.78	93.84	92.41
TrafficDuringWindyDay	98.72	95.04	**99.19**	82.56	98.38	96.89	98.18	79.58	77.68	94.50	93.39
RabbitInNight	98.84	99.75	99.22	91.87	**99.85**	99.51	99.79	82.78	81.41	99.80	99.74
BewareOfTrains	95.19	97.19	97.05	90.75	**99.28**	98.13	97.94	94.10	93.57	97.88	97.28
Average	96.22	94.55	95.17	89.61	**99.43**	98.16	95.77	86.90	83.80	96.68	96.14

TABLE 5. Comparative analysis of false positive rate.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	6.82	1.03	0.33	16.31	0.02	**0.01**	0.47	27.31	29.51	0.44	0.51
Sofa	0.77	0.34	1.79	0.56	**0.02**	0.14	0.50	1.78	2.20	0.07	0.09
WinterDriveAway	5.38	1.69	0.45	9.00	**0.04**	0.13	0.71	11.92	41.59	2.50	2.73
Bootstraping	1.40	1.82	3.27	8.33	**0.22**	1.32	0.25	3.17	3.03	1.83	2.41
WonderingStudents	0.70	0.50	0.08	0.99	**0.08**	0.80	0.69	1.51	1.92	0.33	0.41
Boulevard	1.10	8.63	1.93	14.42	0.89	0.20	0.68	8.32	**0.00**	1.96	2.68
Camouflage	8.45	4.84	**1.85**	24.98	2.64	5.59	11.64	88.69	88.91	6.84	8.80
TimeOfDay	26.58	0.02	13.35	2.33	0.01	0.95	0.34	1.70	22.38	**0.00**	**0.00**
Backdoor	0.15	0.74	0.10	5.00	**0.00**	0.10	1.45	6.74	7.72	0.14	0.24
Bunglaows	2.37	2.18	2.02	1.07	**1.02**	1.75	3.59	4.66	4.91	2.18	2.18
BusStation	1.12	0.93	0.36	3.05	**0.03**	0.21	0.18	2.51	2.83	0.18	0.22
Lobby	5.04	2.56	2.54	10.94	**0.02**	0.15	3.45	10.43	9.45	3.75	4.37
LightSwitch	26.45	85.44	92.01	83.57	**1.17**	26.02	94.83	40.31	79.42	63.49	70.02
Canoe	0.75	6.95	0.17	12.94	1.37	0.08	**0.01**	16.52	17.59	0.63	0.88
Fountain01	0.97	2.34	1.05	4.08	0.54	0.29	**0.08**	9.27	9.67	0.53	0.71
Fountain02	0.05	0.22	0.04	3.74	0.05	0.03	0.01	8.03	**0.00**	0.03	0.06
Overpass	0.59	1.65	1.81	5.26	0.25	0.14	0.04	6.25	**0.00**	0.21	0.30
Campus	0.94	6.62	0.94	13.31	1.01	0.51	**0.36**	7.54	8.60	0.75	1.04
Fountain	1.58	0.80	0.25	7.61	**0.12**	0.39	0.44	5.20	8.08	0.52	0.73
WaterSurface	0.83	0.32	**0.03**	1.87	0.05	0.13	0.46	6.97	10.60	0.05	0.09
WavingTrees	3.95	8.77	**1.54**	13.93	1.78	3.85	2.00	49.99	51.38	2.03	3.04
Carnegie Mellon	0.22	2.13	0.23	1.20	**0.07**	0.16	0.15	2.44	2.86	0.44	0.62
ShoppingMall	2.71	0.65	0.56	5.35	**0.30**	1.19	1.30	3.21	3.36	0.53	0.74
TrainInTunnel	5.11	1.67	2.07	5.62	**0.80**	1.50	2.00	4.28	4.44	1.18	1.43
SideWalk	1.44	11.62	2.30	6.26	1.47	**1.00**	1.17	23.16	23.41	1.10	1.48
Traffic	2.47	6.12	5.07	9.53	**1.18**	2.54	3.05	10.84	13.83	2.77	3.11
Curtain	0.48	0.81	0.71	1.27	0.11	0.20	0.23	9.17	20.86	**0.08**	0.10
Escalator	3.47	5.45	15.38	11.27	**0.50**	4.04	0.70	7.24	7.92	0.55	0.78
Hall	3.48	1.42	0.39	4.40	**0.32**	0.67	1.08	3.45	4.46	0.47	0.61
BigTrucks	2.19	3.65	2.37	19.76	**0.39**	1.19	3.53	6.18	6.22	6.16	7.59
TrafficDuringWindyDay	1.28	4.96	**0.81**	17.44	1.62	3.11	1.82	20.42	22.32	5.50	6.61
RabbitInNight	1.16	0.25	0.78	8.13	**0.15**	0.49	0.21	17.22	18.59	0.20	0.26
BewareOfTrains	4.81	2.81	2.95	9.25	**0.72**	1.87	2.06	5.90	6.43	2.12	2.72
Average	3.78	5.45	4.83	10.39	**0.57**	1.84	4.23	13.10	16.20	3.32	3.86

TABLE 6. Comparative analysis of false negative rate.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	85.73	69.88	87.19	32.30	96.00	94.42	94.70	9.74	**8.89**	78.32	74.94
Sofa	**17.14**	70.43	66.85	30.15	87.52	67.63	68.24	42.38	41.39	66.25	64.52
WinterDriveAway	39.31	41.82	55.67	24.70	91.86	76.79	52.39	**17.28**	50.18	48.25	46.35
Bootstrapping	68.73	49.23	43.70	34.79	82.87	**29.80**	82.76	39.03	40.47	75.76	64.74
WonderingStudents	**4.80**	14.61	25.22	10.14	50.16	9.62	5.42	8.14	7.79	15.20	13.04
Boulevard	16.74	36.39	51.73	**16.43**	78.22	42.72	42.56	21.00	100.00	48.89	40.59
Camouflage	19.93	5.71	4.53	2.17	37.53	**0.00**	9.22	15.49	15.88	13.08	10.94
TimeOfDay	**4.73**	64.46	9.94	81.64	74.20	8.21	65.09	32.20	30.04	89.50	80.53
Backdoor	**10.75**	38.37	50.43	22.05	69.31	17.87	21.45	23.32	23.00	48.42	40.95
Bunglaows	**3.14**	29.73	36.35	26.93	65.11	9.16	11.02	24.33	24.00	32.25	24.45
BusStation	**15.37**	56.59	63.26	24.06	90.55	61.46	64.87	24.66	23.96	56.52	50.15
Lobby	**17.10**	63.51	75.98	25.80	82.03	61.93	68.04	43.44	80.79	78.79	73.94
LightSwitch	**0.73**	65.82	5.83	10.15	83.51	5.58	24.31	20.72	64.74	62.66	54.97
Canoe	23.71	30.00	43.47	**13.18**	90.25	63.64	78.20	14.17	13.65	54.62	34.24
Fountain01	29.10	57.13	69.70	38.02	76.39	35.88	47.47	26.42	**25.54**	68.27	66.23
Fountain02	16.09	42.45	68.58	19.84	71.12	27.27	34.30	**9.02**	100.00	59.19	45.12
Overpass	**17.23**	37.39	46.52	14.83	85.14	56.98	53.86	18.58	100.00	45.95	39.62
Campus	**6.06**	35.17	54.38	12.81	53.14	9.08	38.62	17.74	16.95	63.47	51.47
Fountain	9.95	34.35	45.29	20.15	60.29	25.61	34.73	9.54	**8.60**	56.94	52.72
WaterSurface	**4.09**	20.87	23.55	11.43	82.71	68.18	30.54	7.53	6.70	23.67	23.34
WavingTrees	2.69	3.11	20.90	0.65	32.06	**0.00**	25.02	20.95	20.63	50.70	3.68
Carnegie Mellon	**4.51**	29.89	55.26	6.48	66.97	27.65	18.60	9.37	9.06	32.67	24.27
ShoppingMall	**11.01**	50.17	55.31	23.98	65.21	26.69	38.36	24.29	23.81	56.79	50.75
TrainInTunnel	**13.81**	47.17	63.17	22.11	89.64	75.99	46.70	23.47	22.09	49.76	45.54
SideWalk	52.52	61.80	69.33	44.79	87.23	70.40	89.74	42.54	**41.94**	72.37	67.77
Traffic	21.79	27.22	38.73	**17.69**	62.76	18.60	34.34	22.71	24.14	37.18	32.76
Curtain	8.85	38.26	44.05	13.83	80.47	63.84	45.55	5.18	**3.82**	40.47	37.10
Escalator	17.50	29.94	30.47	18.84	81.23	21.25	78.63	17.82	**16.41**	56.31	49.07
Hall	**10.50**	43.32	59.24	15.45	69.18	53.18	52.37	20.27	18.51	50.70	47.02
BigTrucks	57.71	40.13	60.42	32.91	80.62	63.89	41.67	**20.42**	20.99	47.13	41.44
TrafficDuringWindyDay	22.73	44.98	53.15	26.13	55.64	**13.83**	34.39	22.43	22.23	52.59	39.58
RabbitInNight	17.60	33.71	39.23	15.70	59.48	14.45	22.90	1.84	**1.62**	40.53	37.49
BewareOfTrains	46.02	63.55	78.03	44.83	85.24	58.61	62.02	41.92	**41.38**	68.30	63.27
Average	**21.14**	41.73	48.35	22.88	73.44	38.79	46.00	21.15	31.79	52.77	45.23

TABLE 7. Comparative analysis of recall.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	14.27	30.12	12.81	67.70	4.00	5.58	5.30	90.26	**91.11**	21.68	25.06
Sofa	**82.86**	29.57	33.15	69.85	12.48	32.37	31.76	57.62	58.61	33.75	35.48
WinterDriveAway	60.69	58.18	44.33	75.30	8.14	23.21	47.61	**82.72**	49.82	51.75	53.65
Bootstrapping	31.27	50.77	56.30	65.21	17.13	**70.20**	17.24	60.97	59.53	24.24	35.26
WonderingStudents	**95.20**	85.39	74.78	89.86	49.84	90.38	94.58	91.86	92.21	84.80	86.96
Boulevard	83.26	63.61	48.27	**83.57**	21.78	57.28	57.44	79.00	1.00	51.11	59.41
Camouflage	80.07	94.29	95.47	97.83	62.47	**100.00**	90.78	84.51	84.12	86.92	89.06
TimeOfDay	**95.27**	35.54	90.06	18.36	25.80	91.79	34.91	67.80	69.96	10.50	19.47
Backdoor	**89.25**	61.63	49.57	77.95	30.69	82.13	78.55	76.68	77.00	51.58	59.05
Bunglaows	**96.86**	70.27	63.65	73.07	34.89	90.84	88.98	75.67	76.00	67.75	75.55
BusStation	**84.63**	43.41	36.74	75.94	9.45	38.54	35.13	75.34	76.04	43.48	49.85
Lobby	**82.90**	36.49	24.02	74.20	17.97	38.07	31.96	56.56	19.21	21.21	26.06
LightSwitch	**99.27**	34.18	94.17	89.85	16.49	94.42	75.69	79.28	35.26	37.34	45.03
Canoe	76.29	70.00	56.53	**86.82**	9.75	36.36	21.80	85.83	86.35	45.38	65.76
Fountain01	70.90	42.87	30.30	61.98	23.61	64.12	52.53	73.58	**74.46**	31.73	33.77
Fountain02	83.91	57.55	31.42	80.16	28.88	72.73	65.70	**90.98**	1.00	40.81	54.88
Overpass	**82.77**	62.61	53.48	85.17	14.86	43.02	46.14	81.42	1.00	54.05	60.38
Campus	**93.94**	64.83	45.62	87.19	46.86	90.92	61.38	82.26	83.05	36.53	48.53
Fountain	**90.05**	65.65	54.71	79.85	39.71	74.39	65.27	90.46	91.40	43.06	47.28
WaterSurface	**95.91**	79.13	76.45	88.57	17.29	31.82	69.46	92.47	93.30	76.33	76.66
WavingTrees	97.31	96.89	79.10	99.35	67.94	**100.00**	74.98	79.05	79.37	49.30	96.32
Carnegie Mellon	**95.49**	70.11	44.74	93.52	33.03	72.35	81.40	90.63	90.94	67.33	75.73
ShoppingMall	**88.99**	49.83	44.69	76.02	34.79	73.31	61.64	75.71	76.19	43.21	49.25
TrainInTunnel	**86.19**	52.83	36.83	77.89	10.36	24.01	53.30	76.53	77.91	50.24	54.46
SideWalk	47.48	38.20	30.67	55.21	12.77	29.60	10.26	57.46	**58.06**	27.63	32.23
Traffic	78.21	72.78	61.27	**82.31**	37.24	81.40	65.66	77.29	75.86	62.82	67.24
Curtain	91.15	61.74	55.95	86.17	19.53	36.16	54.45	94.82	**96.18**	59.53	62.90
Escalator	82.50	70.06	69.53	81.16	18.77	78.75	21.37	82.18	**83.59**	43.69	50.93
Hall	**89.50**	56.68	40.76	84.55	30.82	46.82	47.63	79.73	81.49	49.30	52.98
BigTrucks	42.29	59.87	39.58	67.09	19.38	36.11	58.33	**79.58**	79.01	52.87	58.56
TrafficDuringWindyDay	77.27	55.02	46.85	73.87	44.36	**86.17**	65.61	77.57	77.77	47.41	60.42
RabbitInNight	82.40	66.29	60.77	84.30	40.52	85.55	77.10	98.16	**98.38**	59.47	62.51
BewareOfTrains	53.98	36.45	21.97	55.17	14.76	41.39	37.98	58.08	**58.62**	31.70	36.73
Average	**78.86**	58.27	51.65	77.12	26.56	61.21	54.00	78.85	68.30	47.23	54.77

TABLE 8. Comparative analysis of precision.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	25.54	82.76	86.53	40.48	97.74	**98.54**	49.38	35.13	33.60	88.98	88.92
Sofa	88.00	85.54	56.35	89.52	**97.84**	94.11	77.68	68.88	64.56	97.08	96.53
WinterDriveAway	14.39	33.90	59.30	11.09	**75.21**	72.61	37.23	9.38	1.75	23.57	22.63
Bootstrapping	79.18	82.59	74.52	57.04	**92.98**	90.01	92.13	76.56	76.90	69.23	71.31
WonderingStudents	74.50	78.83	**95.41**	66.36	93.32	71.07	74.81	56.85	50.95	84.91	82.00
Boulevard	82.01	30.78	60.19	25.92	59.58	**94.59**	77.43	36.42	1.00	61.14	57.26
Camouflage	91.81	95.84	**98.39**	82.24	96.55	95.48	90.22	52.98	52.80	93.76	92.29
TimeOfDay	22.49	99.42	35.32	39.00	99.46	88.71	89.17	76.35	20.19	**100.00**	**100.00**
Backdoor	94.43	70.85	93.80	31.36	**99.51**	96.07	55.77	24.99	22.61	91.26	87.75
Bunglaows	75.97	71.45	70.92	**84.07**	72.68	80.09	61.00	55.70	54.52	70.67	72.84
BusStation	83.51	75.75	87.14	62.45	**95.06**	92.52	90.52	66.78	64.25	94.23	93.85
Lobby	27.86	25.11	18.18	13.74	**95.48**	85.68	18.22	11.30	4.56	11.72	12.30
LightSwitch	42.61	7.33	16.84	17.54	**73.66**	41.79	13.64	28.01	8.07	10.42	11.29
Canoe	84.59	35.19	94.64	26.56	27.78	96.02	**98.24**	21.88	20.93	79.51	80.14
Fountain01	10.46	2.85	4.41	2.37	6.57	26.20	**41.41**	1.25	1.22	8.69	7.06
Fountain02	86.37	48.55	73.69	7.27	66.09	90.35	**96.75**	3.98	1.00	81.58	77.46
Overpass	72.92	41.95	36.07	23.60	52.73	85.21	**93.59**	19.93	1.00	82.90	79.58
Campus	72.18	20.27	55.77	14.53	54.63	**82.34**	81.44	22.06	20.03	55.91	54.87
Fountain	66.93	74.44	88.53	27.15	**92.38**	87.28	84.29	38.21	28.67	74.50	69.81
WaterSurface	90.91	95.55	**99.59**	80.48	96.57	95.50	92.97	53.58	43.37	99.22	98.66
WavingTrees	91.58	82.96	**95.78**	75.88	94.40	91.97	94.31	41.08	40.52	91.47	93.32
Carnegie Mellon	89.43	39.59	79.76	60.79	90.10	89.93	**91.36**	42.56	38.79	75.10	70.93
ShoppingMall	68.17	83.31	83.89	48.09	**88.25**	80.12	75.38	60.59	59.69	84.20	81.29
TrainInTunnel	43.51	59.15	44.81	38.76	37.05	42.26	56.02	44.96	44.52	**66.05**	63.47
SideWalk	**65.03**	15.65	42.95	33.25	32.95	62.51	13.72	12.28	12.28	58.58	55.06
Traffic	72.25	49.46	49.86	41.55	72.28	**72.51**	42.01	36.99	31.10	65.14	64.04
Curtain	94.94	88.16	88.54	86.98	94.46	94.76	95.96	50.40	31.20	**98.68**	98.48
Escalator	52.00	36.93	17.08	24.70	63.03	47.04	57.60	34.08	32.47	**78.24**	74.89
Hall	63.68	73.14	87.58	56.67	86.85	82.67	75.46	61.14	55.42	**87.63**	85.53
BigTrucks	64.26	60.44	60.83	24.03	**82.07**	73.82	58.21	54.54	54.21	44.44	41.81
TrafficDuringWindyDay	**64.21**	24.85	63.30	11.21	45.02	45.23	52.06	10.17	9.41	20.44	21.42
RabbitInNight	47.17	77.01	49.62	11.53	77.24	68.57	**79.07**	6.68	6.24	78.51	75.49
BewareOfTrains	40.53	44.05	31.11	26.59	55.43	**57.34**	52.03	37.43	35.65	47.59	45.10
Average	64.95	57.38	63.66	40.69	74.70	**77.97**	68.46	37.97	31.01	68.95	67.50

TABLE 9. Comparative analysis of F_1-score.

Video	BCCPDI	Pfinder	FuzzyIntegral	SOBS	IABMM	MultiLayer	PBAS	FPCP	GoDec	VibeGray	VibeRGB
Parking	18.31	44.17	22.31	50.67	7.69	10.56	9.58	**50.57**	49.09	34.86	39.10
Sofa	**85.36**	43.94	41.74	78.47	22.13	48.17	45.09	62.75	61.44	50.09	51.89
WinterDriveAway	23.26	42.84	**50.73**	19.34	14.69	35.18	41.78	16.84	3.39	32.39	31.83
Bootstrapping	44.84	62.89	64.14	60.85	28.93	**78.88**	29.04	67.88	67.11	35.90	47.19
WonderingStudents	**83.59**	81.98	83.85	76.34	64.97	79.57	83.54	70.23	65.63	84.86	84.41
Boulevard	**82.63**	41.49	53.58	39.57	31.90	71.35	65.95	49.86	0.50	55.67	58.31
Camouflage	85.54	95.06	96.91	89.36	75.86	**97.69**	90.50	65.13	64.88	90.21	90.65
TimeOfDay	36.39	52.36	50.74	24.96	40.97	**90.23**	50.17	71.82	31.34	19.01	32.60
Backdoor	**91.77**	65.92	64.87	44.72	46.92	88.55	65.23	37.69	34.95	65.90	70.60
Bunglaows	**85.15**	70.86	67.09	78.18	47.14	85.13	72.38	64.17	63.50	69.18	74.17
BusStation	**84.07**	55.20	51.69	68.53	17.19	54.42	50.62	70.80	69.65	59.51	65.12
Lobby	41.70	29.75	20.69	23.19	30.25	**52.71**	23.21	18.83	7.37	15.10	16.71
LightSwitch	**59.63**	12.08	28.57	29.35	26.95	57.94	23.11	41.40	13.14	16.30	18.05
Canoe	**80.23**	46.84	70.78	40.68	14.43	52.75	35.68	34.87	33.69	57.78	72.24
Fountain01	18.23	5.34	7.69	4.57	10.27	37.20	**46.31**	2.46	2.39	13.64	11.68
Fountain02	**85.12**	52.67	44.05	13.33	40.20	80.59	78.26	7.63	0.50	54.40	64.24
Overpass	**77.53**	50.24	43.08	36.96	23.18	57.17	61.81	32.02	0.50	65.44	68.66
Campus	81.63	30.88	50.19	24.91	50.45	**86.42**	70.00	34.80	32.28	44.19	51.51
Fountain	76.79	69.77	67.63	40.53	55.54	**80.32**	73.57	53.73	43.64	54.57	56.38
WaterSurface	**93.35**	86.57	86.50	84.33	29.32	47.74	79.51	67.85	59.22	86.28	86.28
WavingTrees	94.36	89.39	86.64	86.04	79.01	**95.82**	83.54	54.07	53.65	64.07	94.80
Carnegie Mellon	**92.36**	50.60	57.33	73.68	48.34	80.19	86.09	57.92	54.38	71.00	73.25
ShoppingMall	**77.20**	62.36	58.31	58.91	49.91	76.56	67.82	67.31	66.94	57.11	61.34
TrainInTunnel	**57.83**	55.81	40.43	51.76	16.19	30.62	54.62	56.65	56.66	57.07	58.62
SideWalk	**54.89**	22.21	35.78	41.50	18.41	40.18	11.74	20.23	20.27	37.55	40.66
Traffic	75.11	58.89	54.98	55.23	49.15	**76.70**	51.24	50.04	44.12	63.96	65.60
Curtain	**93.01**	72.62	68.57	86.58	32.36	52.35	69.48	65.82	47.11	74.26	76.77
Escalator	**63.79**	48.37	27.42	37.87	28.92	58.90	31.18	48.18	46.78	56.07	60.63
Hall	**74.42**	63.86	55.63	67.86	45.50	59.79	58.40	69.21	65.98	63.10	65.43
BigTrucks	51.01	60.16	47.96	35.38	31.35	48.50	58.27	**64.72**	64.30	48.29	48.79
TrafficDuringWindyDay	**70.13**	34.24	53.85	19.47	44.68	59.33	58.05	17.99	16.79	28.56	31.62
RabbitInNight	60.00	71.25	54.63	20.29	53.16	76.12	**78.07**	12.52	11.73	67.67	68.39
BewareOfTrains	46.30	39.89	25.76	35.89	23.32	**48.08**	43.91	45.52	44.34	38.05	40.48
Average	**68.05**	53.65	52.55	48.46	36.34	63.51	55.99	47.02	39.31	52.49	56.91

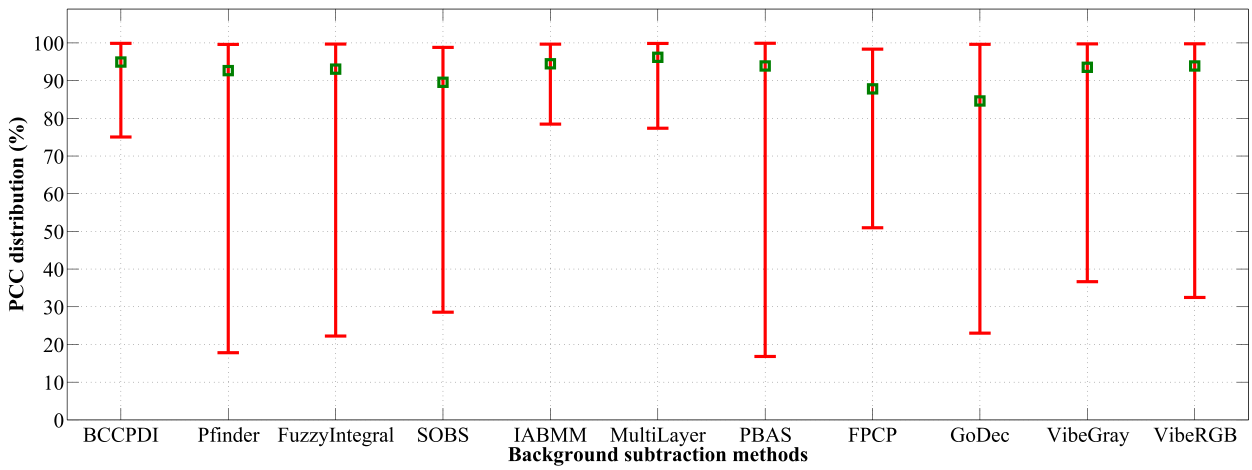

FIGURE 3. PCC Distribution (%) across various methods.

4) ANALYSIS ON CAMOUFLAGE

The attire similarity between the foreground and background can be observed in the *Camouflage* and *Curtain* sequence. All methods except `IABMM`, `FPCP`, `GODec` possess comparatively better output. Complementary cues, *i.e*, texture features along with color cues need to be incorporated to tackle this disguise issue. In addition, the morphological processing and other low pass filtering can be applied as a post improvisation module to minimize the camouflage gap.

5) ANALYSIS ON GRADUAL ILLUMINATION VARIATION

The varying sunlight illumination, over the time, can be seen in the `TimeOfDay` sequence. `Multilayer` is the only method that produces acceptable results. Recursive models often fail to tackle such eventual variations because their underlying model parameters are skewed towards the long past data. On the other hand, non-recursive methods efficiently handle the problem at the cost of high memory overhead in terms of a finite buffer at each pixel location.

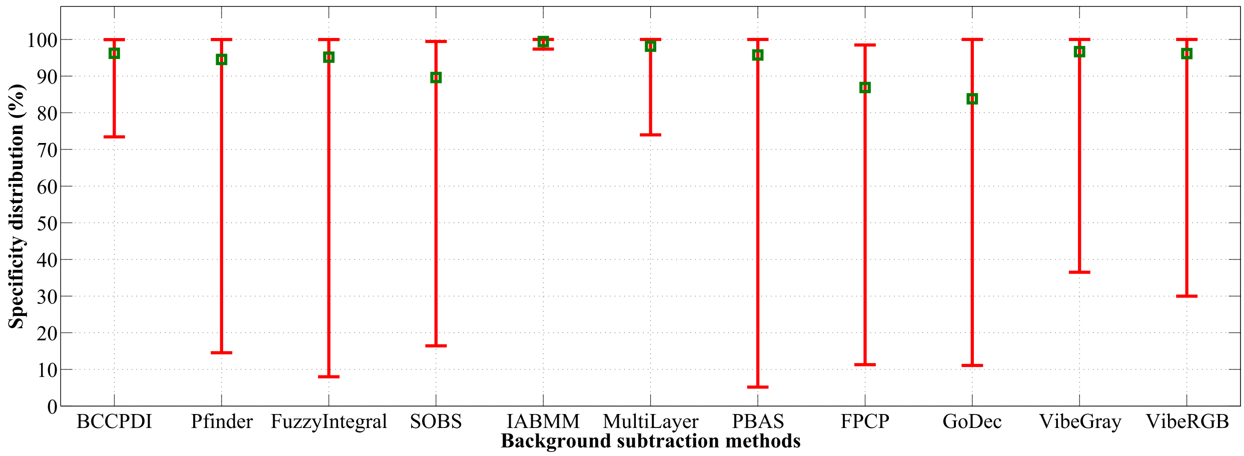

FIGURE 4. **Specificity Distribution (%) across various methods.**

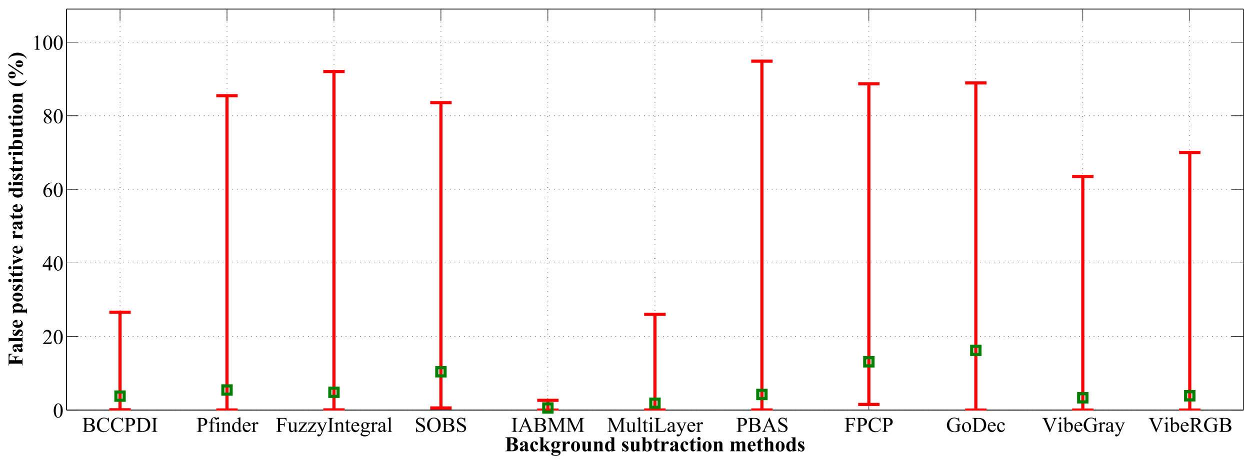

FIGURE 5. **False positive rate Distribution (%) across various methods.**

6) ANALYSIS ON SUDDEN ILLUMINATION VARIATION

The rapid variation in illumination can be observed in the *Lobby* and *LightSwitch* sequence. To the best of our knowledge, the literature still lags any immediate foolproof solution to tackle such rapid variation. `MultiLayer` and `BCCPDI` produce comparatively better results. Such rapid variation completely alters the color and intensity characteristics of the underlying scene. One time-consuming yet reliable solution is to re-initialize the model as soon as such rapid variation is observed. The usual background update strategy also adapts the changed pixel values in the model with few successive frames.

7) ANALYSIS ON SHADOW

Shadow effect can be visualized in the *Backdoor, Bungalow,* and *BusStation* sequences. `BCCPDI` and `Multilayer` efficiently suppress the shadow illumination as compared to other algorithms. Shadow is the scaled down value of illumination. Methods based on *RGB* or gray color space miss-classify shadow as foreground. Gradient or texture features along with invariant color models are suitable candidates to counter this phenomenon.

8) ANALYSIS ON UNINTERESTING BACKGROUND OSCILLATION

We simulate eight sequences (four from CDW, three from I2R, one from Wallflower) that demonstrate various real world instances of background oscillation. `BCCPDI` has the most promising detection rate over others. Unimodal methods fail to incorporate dynamic background in the model. Multi-modal systems usually assign equal number of classes, and therefore fail in situations, where the waving periodicity differs across the scene. The obvious strategy is to learn sample variation of pixel sequence at each location to determine the oscillation periodicity. Then, a suitable clustering method can distribute the input sequence into the required number of classes.

B. ANALYSIS FROM METRIC PERSPECTIVE

The above analyses are useful in evaluating an algorithm against an underlying challenge. However, a desired algorithm should be capable of countering many challenges at a time. In our simulation, we have considered those image sequences that depict more than one challenge simultaneously. The overall results distribution of each

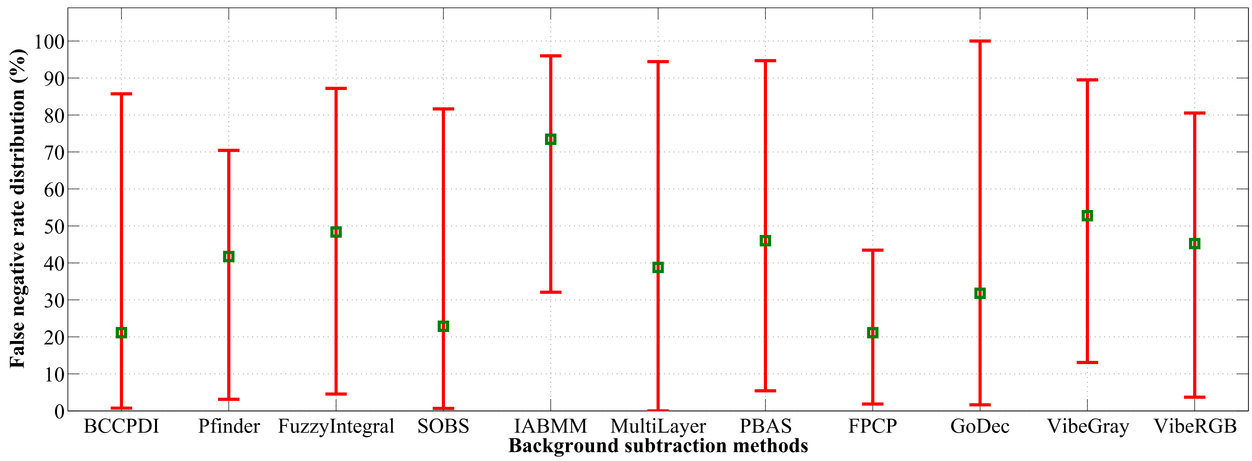

FIGURE 6. False negative rate Distribution (%) across various methods.

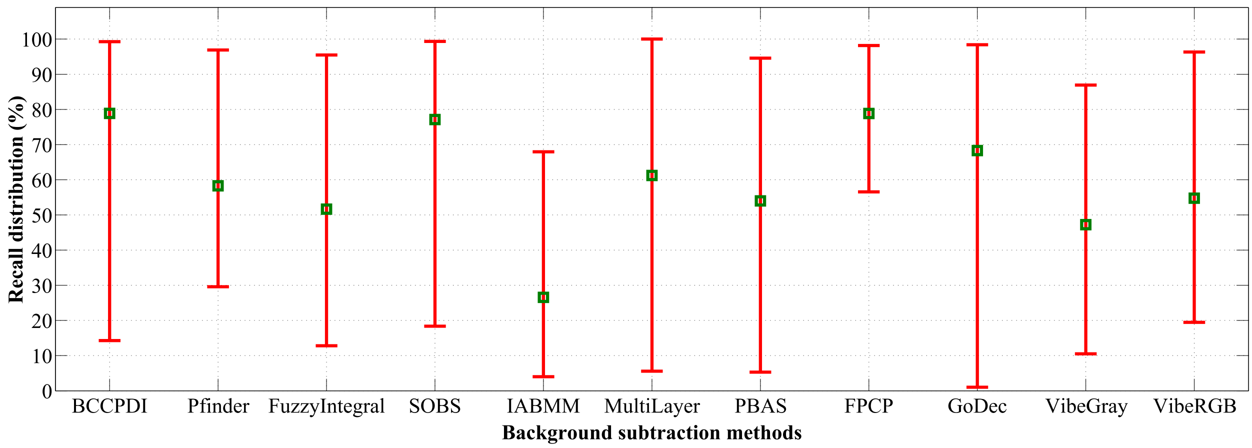

FIGURE 7. Recall Distribution (%) across various methods.

simulated algorithm on the image sequences, in terms of the seven evaluation metrics, are summarized below.

1) Correct classification rate: PCC measures the percentage of correctly detected samples over the entire sample space. Almost all methods possess high PCC as shown in Figure 3.
2) Specificity rate: Specificity measures the proportion of correctly detected background pixels out of total background pixels present in the ground-truth. It can be realized from Figure 4 that all methods except `FPCP` and `GoDec` have an average of more than 90% specificity rate. The variation in results across the image sequences is minimum (the average specificity rate is maximum and reliable) in the case of `IABMM` followed by `MultiLayer` and `BCCPDI`. Usually, the number of foreground pixels in any frame is very less as compared to that of the background pixels. Thereby, the ratio of true negatives to the total negatives will be obviously very high unless the underlying scene is highly multi-modal or the foreground density is very high. Therefore, this metric is not very much well suited to rank the simulated methods.
3) False positive rate: It measures the proportion of incorrectly labeled foreground pixels out of the total background samples available in the ground-truth. False positives in the case of background subtraction occur due to an oscillating background, bootstrapping movement, relocating stationary objects, varying illumination, shadow impression *etc*. The distribution of false positive rate is plotted in Figure 5. All methods except `FPCP` and `GoDec` have comparatively low false positive rate. Again, the variation in the distribution is minimal in case of `IABMM`, `MultiLayer` and `BCCPDI`.
4) False negative rate: It measures the percentage of incorrectly labeled background pixels out of the total foreground samples in the binary ground-truth. Almost all methods have a high false negative rate that can be visualized in Figure 6. The disguise issue owing to attire similarity between the foreground and rearward background yields significant false negatives. Another major factor is the selection of deviation threshold that acts as a decision boundary between the foreground and background regions; a large threshold may

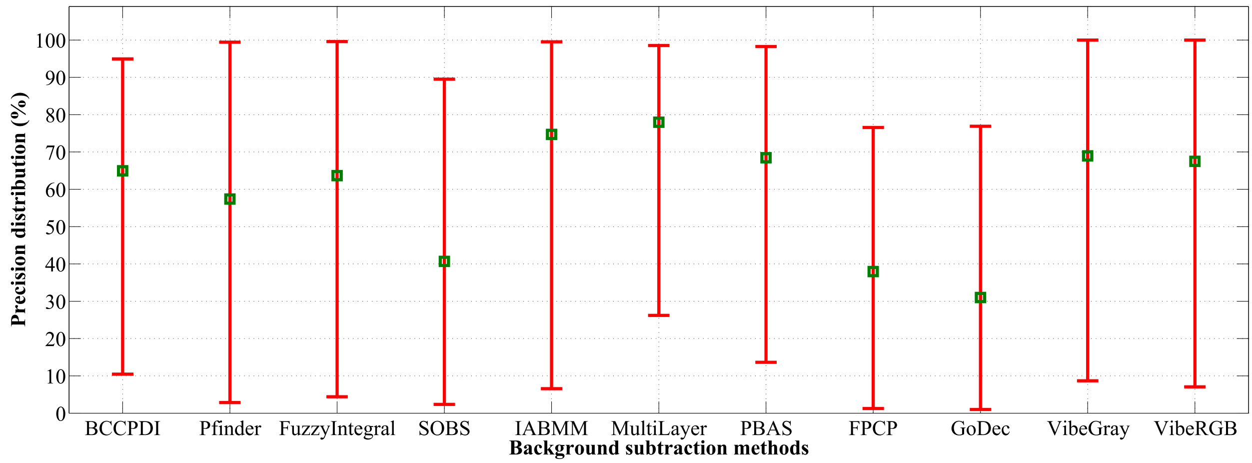

FIGURE 8. Precision Distribution (%) across various methods.

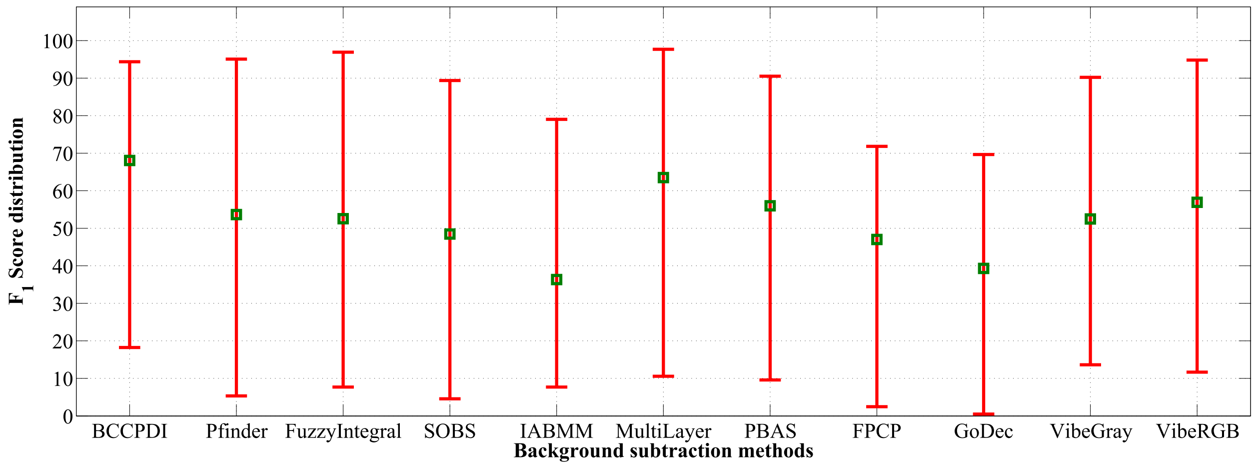

FIGURE 9. F_1 Score Distribution (%) across various methods.

incorrectly include the foreground pixels in the background regions resulting in false negatives.

5) Recall rate: Recall measures the percentage of true foreground pixels detected by any algorithm over the actual collection of foregrounds in the ground-truth. The Recall distribution is depicted in Figure 7. `FPCP`, `BCCPDI`, and `SOBS` have comparatively better results over their counterparts. On the contrary, the average recall rate in case of `IABMM` is even less than 50%.
6) Precision rate: It measures the percentage of correctly labeled foreground samples out of all positive samples detected by any algorithm. The corresponding distribution result is shown in Figure 8. `MultiLayer` and `IABMM` have maximum recall rate, whereas `FPCP`, `GoDec`, `SOBS` possess the least among others.
7) F_1 Score distribution: Neither Recall nor Precision alone may accurately measure the efficiency of the simulated algorithms; rather their combination is a better choice to select the superior methods. The distribution of F_1 Score, plotted in Figure 9, ranks `BCCPDI` and `MultiLayer` as the two best methods.

C. PARAMETER SELECTION

One major concern in background subtraction is the choice of appropriate parameters. The effect of various parameters applied across different phases of background subtraction is enumerated below.

1) Accurate modeling of a scene is directly related to the number of frames (say M) adapted to create the background model. The parameter M should be varied depending on the scene structure. Higher is the foreground density or bootstrapping movements, the more is the number of initialization frames required to capture all background locations precisely. Furthermore, the periodicity of waving background at all locations as well as the camera oscillation need to be learned with sufficient training frames that are again directly relative to the number of initialization frames M.
2) Foreground extraction phase requires detailed knowledge of the size, speed, halt duration of possible mobile objects to formulate an appropriate deviation threshold to separate the foreground pixels.

3) Background maintenance phase requires the scene knowledge to decide (i) the immobile duration threshold beyond which a foreground object will be absorbed in the model, and (ii) the absence duration threshold beyond which an existing background will be removed from the model.
4) The shadow illumination may increase or decease depending on the intensity of source illumination. Accordingly, the shadow removal threshold should be modeled as a function of light illumination rather than a constant value.
5) Another major concern is the temporal buffer length in case of non-recursive modeling that holds the recent pixel history. A small sized buffer may fail to appropriately model a background location. On the contrary, a large-sized buffer may include the long past observation that may result in false negatives together with higher memory overhead.

VI. CONCLUSION

This paper includes a detailed evaluation of various background subtraction framework for detecting objects moving across a scene. The principles adopted by the reported methods for mathematical modeling is taken into consideration to classify their evolution: parametric vs. non-parametric model, unimodal vs. multi-modal background, pixel-based vs. region-based segmentation, recursive paradigm vs. non-recursive architecture *etc*. We have enumerated the possible challenges that come into picture during background subtraction along with their varying mitigation strategies over the years. Some of the state-of-the-art methods are simulated on thirty-four benchmark image sequences, in which each sequence portrays either a single challenge or a number of challenges at a time. A set of seven benchmark evaluation measures is selected to compare the output sequence with the supplied ground-truth. The variation in result distribution across the simulated image sequences gives an idea to select the suitable method depending upon the requirement.

The underlying scene knowledge along with more prior details regarding the foreground movements and available background objects are very much helpful in formulating (1) the absence duration beyond which the existing background class(es) will be removed from the background model, and (2) the appearance interval after which an immobile foreground will be relabeled as background. The varying oscillating pattern across the background (at each model location) has to be learned with sufficient initialization frames to address the uninteresting background movement and camera motion. Complementary texture details need to be incorporated along with invariant color features to tackle the problem with camouflage and shadow illumination. Finite queue (with recently accessed background pixels only) driven non-recursive background modeling is proven effective to cope with slow varying sunlight illumination. No solution to the problem of rapid light illumination variation, to the best of our knowledge, is found in the literature; however, the background update automatically reinitializes the model at the cost of few successive frame delays. It can be observed that the false positive rate is very low for the simulated methods, however, such attempt sometimes substantially increases the false negative rate. It can be realized that the selection of appropriate parameters at each stage of background subtraction demands the prior scene knowledge and more information regarding the possible stationary and non-stationary movements.

REFERENCES

[1] I. Setitra and S. Larabi, "Background subtraction algorithms with post-processing: A review," in *Proc. Int. Conf. Pattern Recognit.*, 2014, pp. 2436–2441.
[2] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.
[3] A. Elgammal, "Background subtraction: Theory and practice," *Synthesis Lectures Comput. Vis.*, vol. 5, no. 1, pp. 1–83, 2014.
[4] K. K. Hati, P. K. Sa, and B. Majhi, "Intensity range based background subtraction for effective object detection," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 759–762, Aug. 2013.
[5] S. Brutzer and B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1937–1944.
[6] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection–A systematic survey," *Recent Patents Comput. Sci.*, vol. 4, no. 3, pp. 147–176, Sep. 2011.
[7] A. H. S. Lai and N. H. C. Yung, "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence," in *Proc. Int. Symp. Circuits Syst.*, vol. 4. May 1998, pp. 241–244.
[8] I. Haritaoglu, D. Harwood, and L. Davis, "W/sup 4/: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, Aug. 2000.
[9] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.
[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 1999, pp. 246–252.
[11] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.
[12] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 2. 2004, pp. 28–31.
[13] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, 2006.
[14] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Video-Based Surveillance Systems*. New York, NY, USA, Springer, 2002, pp. 135–144.
[15] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, Jun. 2005.
[16] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction," *AEU Int. J. Electron. Commun.*, vol. 64, no. 8, pp. 739–747, Aug. 2010.
[17] E. Fernandez-Sanchez, J. Diaz, and E. Ros, "Background subtraction based on color and depth using active sensors," *Sensors*, vol. 13, no. 7, pp. 8895–8915, 2013.
[18] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.
[19] J. M. McHugh, J. Konrad, V. Saligrama, and P.-M. Jodoin, "Foreground-adaptive background subtraction," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 390–393, May 2009.
[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2000.

[21] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 751–767.
[22] M. Piccardi and T. Jan, "Mean-shift background image modelling," in *Proc. Int. Conf. Image Process.*, vol. 5. Oct. 2004, pp. 3399–3402.
[23] A. Elgammal, R. Duraiswami, and L. S. Davis, "Efficient non-parametric adaptive color modeling using fast gauss transform," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Dec. 2001, pp. 563–570.
[24] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun./Jul. 2004, pp. 302–309.
[25] T. Parag, A. Elgammal, and A. Mittal, "A framework for feature selection for background subtraction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. 2006, pp. 1916–1923.
[26] B. P. L. Lo and S. A. Velastin, "Automatic congestion detection system for underground platforms," in *Proc. IEEE Int. Symp. Intell. Multimedia, Video Speech Process.*, May 2001, pp. 158–161.
[27] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
[28] S. Calderara, R. Melli, A. Prati, and R. Cucchiara, "Reliable background suppression for complex scenes," in *Proc. ACM Int. Multimedia Conf. Exhibit.*, 2006, pp. 211–214.
[29] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 1. Sep. 1999, pp. 255–261.
[30] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1. Oct. 2003, pp. 44–50.
[31] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.
[32] H. Wang and D. Suter, "Background subtraction based on a robust consensus method," in *Proc. Int. Conf. Pattern Recognit.*, vol. 1. Aug. 2006, pp. 223–226.
[33] H. Wang and D. Suter, "A consensus-based method for tracking: Modelling background scenario and foreground appearance," *Pattern Recognit.*, vol. 40, no. 3, pp. 1091–1105, Mar. 2007.
[34] T. Bouwmans, "Background subtraction for visual surveillance: A fuzzy approach," *Handbook Soft Comput. Video Surveill.*, 2012, pp. 103–134.
[35] Z. Hongxun and X. De, "Fusing color and texture features for background model," *Fuzzy Syst. Knowl. Discovery* (Lecture Notes in Computer Science), vol. 4223. Berlin, Germany, Springer, 2006, pp. 887–893, doi: 10.1007/11881599_110
[36] F. El Baf, T. Bouwmans, and B. Vachon, "Foreground detection using the Choquet integral," in *Proc. Int. Workshop Image Anal. Multimedia Interact. Services*, May 2008, pp. 187–190.
[37] T. Bouwmans and F. El Baf, "Modeling of dynamic backgrounds by type-2 fuzzy Gaussians mixture models," *MASAUM J. Basic Appl. Sci.*, vol. 1, no. 2, pp. 265–276, 2009.
[38] F. El Baf, T. Bouwmans, and B. Vachon, "Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2009, pp. 60–65.
[39] W. Kim and C. Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 127–130, Mar. 2012.
[40] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht, "Neural network approach to background modeling for video object segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1614–1627, Nov. 2007.
[41] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.
[42] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 179–186, 2010.
[43] P. Rodriguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 69–73.
[44] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, Jun. 2011, pp. 33–40
[45] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
[46] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.
[47] C. Qiu and N. Vaswani, "Real-time robust principal components' pursuit," in *Proc. 48th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep./Oct. 2010, pp. 591–598.
[48] A. Sobral, C. G. Baker, T. Bouwmans, and E.-H. Zahzah, "Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction," in *Proc. Int. Conf. Image Anal. Recognit.*, Oct. 2014, pp. 94–103.
[49] S. Javed, S. H. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "OR-PCA with mrf for robust foreground detection in highly dynamic backgrounds," in *Proc. Asian Conf. Comput. Vis.*, Nov. 2014, pp. 284–299.
[50] L. Li, P. Wang, Q. Hu, and S. Cai, "Efficient background modeling based on sparse representation and outlier iterative removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 2, pp. 278–289, Feb. 2016.
[51] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. Int. Conf. Comput. Vis.*, vol. 99, 1999, pp. 1–19.
[52] W. Zhang, X. Z. Fang, and X. Yang, "Moving cast shadows detection based on ratio edge," in *Proc. Int. Conf. Pattern Recognit.*, vol. 4. Aug. 2006, pp. 73–76.
[53] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Comput. Vis. Image Understand.*, vol. 80, no. 1, pp. 42–56, Oct. 2000.
[54] H. Wang and D. Suter, "A re-evaluation of mixture of Gaussian background modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 2. Mar. 2005, pp. 1017–1020.
[55] L. Li, W. Huang, I. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proc. ACM Int. Multimedia Conf. Exhibit.*, 2003, pp. 2–10.
[56] I. Huerta, M. Holte, T. Moeslund, and J. Gonzýlez, "Detection and removal of chromatic moving shadows in surveillance scenarios," in *Proc. Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1499–1506.
[57] Q. Zhou and J. K. Aggarwal, "Tracking and classifying moving objects from video," in *Proc. IEEE Workshop Perform. Eval. Tracking Surveill.*, Honolulu, HI, USA, 2001.
[58] D. H. Parks and S. S. Fels, "Evaluation of background subtraction algorithms with post-processing," in *Proc. Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2008, pp. 192–199.
[59] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.
[60] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2012, pp. 1–8.
[61] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 387–394.
[62] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved foreground detection via block-based classifier cascade with probabilistic decision integration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 83–93, Jan. 2013.
[63] F. El Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jun. 2008, pp. 1729–1736.
[64] J. Yao and J.-M. Odobez, "Multi-layer background subtraction based on color and texture," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[65] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38–43.
[66] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

SUMAN KUMAR CHOUDHURY received the M.Tech. degree from the National Institute of Technology Rourkela, India, in 2013. He is currently pursuing the Ph.D. degree in computer vision with the Department of Computer Science and Engineering, National Institute of Technology Rourkela. His research interest includes video surveillance, image processing, and pattern recognition. His excellence in research has brought him laurels from the academia.

PANKAJ K. SA received the Ph.D. degree in Computer Science in 2010. He is currently serving as an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. His research interest includes computer vision, biometrics, visual surveillance, and robotic perception. He has co-authored a number of research articles in various journals, conferences, and book chapters. He has co-investigated some Research and Development projects that are funded by SERB, DRDO-PXE, DeitY, and ISRO. He is the recipient of prestigious awards and honors for his excellence in academics and research. Apart from research and teaching, he conceptualizes and engineers the process of institutional automation.

SAMBIT BAKSHI received the Ph.D. degree in computer science in 2015. He is currently with the Centre for Computer Vision and Pattern Recognition, National Institute of Technology Rourkela, India. He also serves as an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela. He serves as an Associate Editor of *International Journal of Biometrics* (2013). He is a Technical Committee Member of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence. He received the Prestigious Innovative Student Projects Award-2011 from Indian National Academy of Engineering for his master's thesis. He has more than 30 publications in journals, reports, and conferences.

BANSHIDHAR MAJHI is a Professor with the Department of Computer Science and Engineering, National Institute of Technology Rourkela, India. He has successfully executed various Research and Development projects being funded by agencies, such as MHRD, ISRO, DRDO, and DeitY. He has authored hundreds of articles in reputed journals and conferences. His current research interests include image processing, computer vision, biometric security, and pattern recognition. He has been conferred with prestigious awards and honors for his contribution toward scientific research and academic excellence.