This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

Kernelized Fuzzy Modal Variation for Local Change Detection from Video Scenes

Badri Narayan Subudhi, T. Veerakumar, S. Esakkirajan and Ashish Ghosh

***Abstract*—Background subtraction (BGS) is a popular scheme epitomized in the state-of-the-art literature on video processing. In this context, a novel online kernelized fuzzy modal variation based background subtraction scheme for detecting local changes from the sequences of image frames is proposed. In the proposed scheme, the time varying background at different instances of time are modeled using fuzzy set theory. The proposed background subtraction scheme, utilizes the fuzzy modal variation as the cost function for fitting the pixel values of the image frames. The use of kernel based modal variation helps in projecting the pixel values in a higher dimensional space, linearly separating them into object and background classes. The results of the proposed technique is verified on different challenging sequences including dynamic background, camera jitter, noise, blurred scene, etc. The proposed technique is successfully tested over several test sequences with two major databases (all sequences) and it provides better results compared to the twenty one existing state-of-the-art techniques.**

***Index Terms*—Background subtraction, Object detection, Temporal analysis, Fuzzy logic, Modal variation.**

I. Introduction

Visual surveillance is one of the challenging areas of research for multimedia analysis. Recently, the demand for visual surveillance has increased due to the growing importance of public safety at different places. Visual surveillance has been used for long time to monitor different areas such as banks, traffic, hall, corridor, borders, etc. The demand for video surveillance can be well cited in many multimedia applications, including face and gait-based human recognition, activity recognition, visual surveillance, robotics, etc [1]. Visual surveillance is achieved by two different ways: manual and automated. In the manual case, the operator controlling the surveillance system needs proper attention to locate the object of interest from the scene. In the automated case, the programmed device will identify the object of interest by itself.

Any automated video surveillance system may refer to a moving object detection followed by tracking. Detection of a moving object is the process of locating those objects in a video, the movements of which creates a dynamic variation in the video scene [1]. The moving object detection scheme is alternatively termed as foreground segmentation [2]. One of the popular and effective ways of foreground segmentation is background subtraction (BGS) [3], where a binary mask is used to represent the locations of the moving objects in the scene. Hence, a video frame is divided as foreground or background.

The state-of-the-art techniques on visual surveillance is enormous. Several algorithms are developed in the allied area of research. Most of the BGS schemes are found to be effective only for images with significant contrast. An image frame possesses high ambiguity within the pixels in spatial and temporal domains due to its possible multi-valued levels of brightness. Most of the techniques developed/studied in the literature used a hard decision to find if the pixel is foreground/background region type, and hence, most of them do not yield satisfactory solution. In the state-of-the-art techniques for BGS, several algorithms are reported which uses fuzzy set theoretic approach for local change detection. Also, the fuzzy based BGS schemes are limited by under performance in non-linear feature space, where kernel based approaches provide better results. Most of the techniques are offline where testing and background training are carried out in separate stages.

In this article, we put forward a novel online kernelized fuzzy modal variation based background subtraction scheme to detect the local changes from the video scenes. The proposed background subtraction scheme, utilizes fuzzy modal variation as the cost function for fitting the pixel values of the image frames. The time varying background at different instant of time are modeled using fuzzy set theory. The use of kernel based modal variation helps in projecting the pixel values in a higher dimensional feature space and linearly separating them into object and background classes. In the proposed scheme, the number of background types/modes are assumed as unknown. There is no need to fix/specify the number of background types. At each pixel location of the image sequence, two modes are assumed to be present, initially. The proposed technique will add more modes to each pixel location by fitting the pixel values to the kernelized fuzzy modal variation due to different background types. In the target frame, each pixel value is tested with the constructed background model. If it fits, then it is considered as a background else detected as a part of local change.

The proposed BGS technique is validated by testing on two major databases with different challenging sequences: dynamic background, camera jitter, noisy, blurred scene, etc.

Authors would like to acknowledge SERB, India for providing the financial support with grant EMR/2016/002552 and DSTICPS support: DST/ICPS/CLUSTER/Data Science/2018 dt. 07.01.2019

B. N. Subudhi is with the Department of Electrical Engineering, Indian Institute of Technology Jammu, Jagti, Nagrota, Jammu, 181121, Jammu and Kashmir, INDIA, email:- subudhi.badri@gmail.com

T. Veerakumar is with the Department of Electronics and Communication Engineering, National Institute of Technology, Farmagudi, Ponda, 403401, Goa, India, email:- tveerakumar@yahoo.co.in

S. Esakkirajan is with the Instrumentation & Control Engineering, PSG College of Technology, Coimbatore, Tamilnadu, India, email:- rajanesakki@yahoo.com

A. Ghosh is with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata, India, email:- ash@isical.ac.in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

It is observed that the proposed scheme is found to provide better results compared to the twenty one existing state-of-the-art techniques. The performance of the proposed technique is evaluated using different performance evaluation measures.

The organization of the remaining part of this article is as follows. Section II gives a description of the state-of-the-art techniques on BGS schemes. A description on the proposed background training and background subtraction are provided in Section III. Section IV presents the experimental results with discussion and future work. Conclusions are drawn in Section V.

II. State-of-the-art Techniques

In the last two decades several BGS techniques had been developed in the literature of computer vision. One of the primary and simple approaches of foreground segmentation is by frame difference followed by thresholding [4]. However, such an approach has two main deficiencies. Selection of the background frame is random and does not take into account the illumination variation and hence, it may attenuate the noise, environmental changes, and other irrelevant details in the foreground mask. It may not be able to distinguish the variations as static or dynamic, whereas dynamic variation is a major concern. Different linear algebric change detection techniques including Gramian matrix, *Wronskian* function, etc. [5] are proposed in the literature. However, most of these techniques are developed assuming the scene to contain static variation and is unable to deal with the background of non-static variation.

Stauffer and Grimson [6] proposed a BGS scheme where a mixture of Gaussian probability density functions (*pdfs*) are used to represent the multi-valued background of a video. The modeled background parameters are used to detect the foreground from the target image frames. However, such an approach rarely carries spatio-contextual information and may produce holes/ghosts in the scene. A robust BGS scheme that uses a combination of minimum spanning tree based aggregation technique with optical flow estimation is proposed by Chen *et al.* [7]. In this regard, Subudhi *et al.* [3] proposed a spatio-contextual BGS technique, where the temporal modes in a video are modeled using the Gaussian *pdf* and spatial modes are fitted by Wronskian function to construct the background. Each pixel in the target frame is compared with the constructed background to detect the local changes. However, such an approach fails to model the multi-valued background in a scene. So, a modification of the same is achieved by Rout *et al.* [8], where the temporal modes are extended by fitting with mixture of Gaussian *pdfs*. A modification of the Stauffer and Grimson's BGS model is also proposed by Chan *et al.* [9], where the dynamic texture is explored for foreground detection. To deal with dynamic background, a novel BGS scheme which uses Dirichlet process Gaussian mixture models for background modeling and a probabilistic regularization process to incorporate the neighboring pixel information is developed [10]. Jodoin *et al.* [11] proposed a BGS scheme where statistical motion detection scheme is explored in a single frame for background modeling. It may be observed that in video data, the number of pixels for the foreground class are minimum and background class are maximum. Hence, this may give rise to class imbalance in BGS. Zhang *et al.* [12] proposed a BGS, where spatio-temporal over-sampling and selective down-sampling are proposed to compensate the data imbalance.

It can be concluded from the above analysis that, it is difficult to estimate the parameters of the distribution properly, as they are too much parameter dependent. In this regard, several non-parametric BGS techniques, which uses kernel or histogram based approaches for parameter estimation are developed in the literature. One of the popular non-parametric BGS schemes is developed by Elgammal *et al.* [13] where the kernel function is used for Gaussian mixture model (GMM) parameter estimation. Kim *et al.* [14] also proposed a BGS technique where the concept of codebook construction is used for modeling the background from the considered video. This approach is found to be more robust and even effective for sequences with non-static background. Spagnolo *et al.* [15] proposed a non-parametric BGS scheme where the radiometric correlation between the pixels in a considered region of support is used for background modeling and foreground detection. Histogram filtering technique [16] is also used in the literature to construct the background model. Recently, Subudhi *et al.* [17] proposed a non-parametric BGS technique where statistical features extracted from the local region of a frame is stored in few virtual bags for constructing the background model. The parameters computed from each bag is considered to be the tag on the bag. For any pixel on the target frame, majority voting technique is exploited to check if it is foreground/background pixel. The concept of local binary pattern is also explored for the construction of the background model and detecting the local changes [18].

It may be observed from the recent literature that the use of principal component analysis (PCA) gets its popularity for local change detection in background subtraction [19]. Several PCA based schemes are explored/developed by the researchers in this field, including binary PCA (BPCA) [19]. Among them, Bayesian inferring with PCA is found to be very powerful. One of the major challenges in PCA based BGS is to cope with long duration sequences. Seo and Kim [20] proposed a BGS technique where the background model is initiated in a low-dimensional subspace and then updated the recursive on-line PCA (RPCA) scheme periodically. Seidel et al. [21] proposed a BGS scheme where the convex relaxation of the ideal sparsifying function is represented using l_p norm, where the algorithm uses alternating minimization on manifolds. ViBe [22] is another popular non-parametric pixel based classification strategy, used for BGS, where random sampling strategy is used with the spatio-contextual property for the construction of the background model.

Low rank and sparse based BGS mechanisms are recently, becomes popular. In a video, due to the movement of objects, distribution of the regions corresponding to different objects in the scene are structurally sparse. Liu et al. [23] proposed a BGS scheme, where a spatial information in sparse outliers are used to model the foreground and low-rank matrix to background. A new online subspace learning based BGS scheme

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

is recently, proposed by Yong *et al.* [24], where the authors have used a mixture of Gaussians (MoG) distribution to model the foreground of each frame. This model is formulated in a concise probabilistic MAP framework and solved by EM algorithm. A newly proposed BGS scheme that uses a modified signal decomposition technique optimized and augmented with Lagrangian framework is proposed by Minaee and Wang [25]. St-Charles *et al.* [26] proposed a BGS scheme where spatiotemporal binary features and color information are used to detect the local changes in a video, which the authors have prioritized for detecting object motion in camouflaged condition. A BGS scheme that uses mega pixels concept to detect the moving object in a video scene is proposed in [27], where multiple innovative mechanisms are followed by the foreground/background probability estimation process for a different pixel locations. Recently, Jiang *et al.* [28] proposed BGS scheme, where a variable weighing mechanism is used to detect the local changes in a scene. Minimum-weight update policy is used to replace the most inefficient sample and a reward-and-penalty weighting strategy is used to reinforce frequently occurring samples. Use of image and color space reduction is also explored in BGS to compensate the color noise [29].

Recently, the compressive sensing mechanism is gaining more attention for background subtraction. However such approaches are limited due to its computational cost. Liu *et al.* [30] proposed an iterative data residues prunning based spatio-temporal group sparsity recovery technique for BGS. Cao et al. [31] proposed a tensor-based robust PCA approach for background subtraction from compressive measurements, where the video frames are decomposed into backgrounds and foregrounds with spatio-temporal tensor framework. Super-pixel segmentation mechanism is also explored for BGS. A hierarchical superpixel segmentation scheme that uses spanning trees and optical flow is proposed by Chen *et. al.* [32]. De Gregorio and Giordano [33] proposed a BGS scheme based on a weightless neural network called BEWiS. The authors have estimated the background at each pixel location using a weightless neural network designed to learn pixel color frequency. Use of deep neural network for background modeling is also explored in the recently proposed change detection schemes in [34], [35].

It may be concluded from the above analysis that in BGS, illumination changes and structural background changes produce uncertainty for the classification of a pixel in foreground and background. Hence, it gives imprecision in the localization of the object. One of the early works reported in the literature is fuzzy running average for background subtraction [36]. However, such an approach rarely utilizes the fuzzy mechanism for modeling the multi-valued background. Zhang and Xu [37] proposed a fuzzy integral based background subtraction scheme, where texture and color features are fused for detecting moving objects from the background and can handle non-static background in a video scene. El Baf *et al.* [38] also proposed a BGS scheme, where Choquet integral is explored in fuzzy framework for background subtraction and used a fuzzy adaptive model for background update. A spatial coherence based background subtraction scheme using self-organization map is proposed by Maddalena and Petrosino [39]. Kim and Kim [40] proposed a background subtraction scheme where the background motion is modeled with fuzzy set to generate new kind of features called fuzzy color histogram which are found to be effective in identifying the moving objects. Fuzzy background subtraction scheme is also developed by Choquet integral process over a set of pixels to eliminate uncertainties [41]. In this context, Bouwmans and El Baf [42] proposed a Type-2 fuzzy GMM for modeling the non-static background and detecting moving objects from it. This approach is found to be effective in complex background condition.

III. Proposed Background Subtraction Process

Most of the early literature concentrated on non-kernelized BGS schemes. One of the major disadvantages of non-kernelized BGS is that for checking the belongingness of a particular pixel value in different background (unchanged area) or object (changed area) is decided based on some linear classification/separation process [43]. It is also to be noted that as the object is moving from one place to another, the gray levels in the video changes from frame to frame. Generally, in a video, the changes in gray level between the frames is very common and different BGS schemes are found to be effective only for images with significant contrast. Again an image frame possesses high ambiguity within the pixels in spatial and temporal domain due to its possible multi-valued levels of brightness. Use of a hard decision for inclusion of pixels in background/ foreground class in BGS is not expected to yield a satisfactory solution. Again, a *3*-dimension *RGB* color space has a large correlation and high non-linearity. A linear separation process with hard decision based classifier fails to provide the required solutions in this case. It is to be noted that fuzzy set theories are applied to handle uncertainties to a reasonable extent, arising from deficiencies of information available from a situation (the deficiency may result from incomplete, ill-defined, not fully reliable, vague and contradictory information). It justifies to apply the concept of fuzzy set based BGS against the hard decision based BGS. The fuzzy set theoretic approaches always may not be able to provide a better solution in this regard and it fails to handle the problem if the data is non-linear due to ambiguities present in it [44]. This justifies and motivates the use of kernelized fuzzy BGS.

The main essence of kernel based fuzzy modal variation is to map the original d-dimensional feature space to a non-linear high dimensional space i.e., the kernel space, with infinite dimensions [45]. One of the main goals of going to a higher dimensional space is to apply a linear classification in the high dimensional kernel space [45] for easy separation of the data points.

Fig. 1 represents the conceptual representation of the proposed scheme. Let us consider an example of vibrating leaves of a tree. Due to quasi-periodic motion of the leaves, there occurs different background types. In the proposed scheme these background types are modeled with different fuzzy membership values. The proposed scheme follows two stages:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

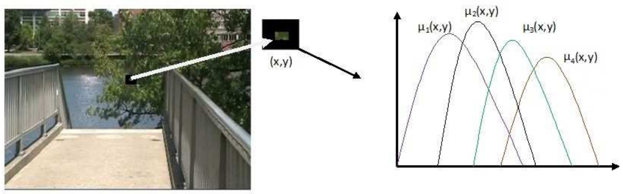

Fig. 1. Background fuzzy modeling

background construction and object separation. It is assumed that the frame instant is the same as that of the time instant. Initially, the input image sequences are divided into two parts: training and testing frames. Here it is assumed that the input video contains N number of frames out of which n is considered as the number of training frames and $N-n$ is the target/testing frames. The pixel at location (x,y) of a frame at i^{th} time instant is represented as $f_i(x,y)$. Each pixel $f_i(x,y)$ in a frame is assumed to be in a three dimensional RGB color space.

A. Background Construction

The proposed background construction process is an automatic one. It does not need any prior information like number of background or background types unlike multiple *pdf* [6] based background construction approaches. Here, at each pixel location (x,y), the cost function of the fuzzy modal variation is fitted. The expression for fuzzy modal variation at any pixel location (x,y) in an image frame at time t, which can be calculated as:

$$J_t(x,y) = \sum_{i=1}^{t}\sum_{j=1}^{m}\{\mu_{ij}^{r}(x,y)\}||\Phi(f_i(x,y)) - \Phi(v_j(x,y))||^2, \quad t \leq n, \quad (1)$$

where $J_t(x,y)$ expresses the kernelized fuzzy modal variation function at location (x,y), computed for frame instant t. m is the number of background types built at location (x,y) and $v_j(x,y)$ is the mode corresponding to the j^{th} background type. The term $\mu_{ij}(x,y)$ represents the fuzzy membership value for a pixel belonging to the j^{th} background type at location (x,y). t represents the time instant and r is the fuzzifier. Φ represents the kernel function used for non-linear mapping of *d*-dimensional space to kernel space. Eq (1) represents the cost function of kernelized fuzzy clustering. Eq (1), expresses the *kernelized fuzzy modal variation* over the pixel values.

The concept of kernel function is inspired from the Mercer's kernel theorem. Mercer's theorem suggests that a distance $||\Phi(x_i) - \Phi(v_j)||^2$ in the kernel space cannot be computed exactly as that of the *d*-dimensional Euclidean space, due to inclusion of kernel function. The kernelized fuzzy method takes the advantage of the dot products in the kernel space and expresses the distance function $||\Phi(f_i(x,y)) - \Phi(v_j(x,y))||^2$ that can be expressed as

$$||\Phi(f_i(x,y)) - \Phi(v_j(x,y))||^2 = K(f_i(x,y), f_i(x,y)) + K(v_j(x,y), v_j(x,y)) - 2K(f_i(x,y), v_j(x,y)), \quad (2)$$

where $K(.)$ represents the dot product. Considering Gaussian kernel function

$$K(f_i(x,y), v_j(x,y)) = exp\{-||f_i(x,y) - v_j(x,y)||^2/\sigma^2\}; \quad (3)$$

the kernelized fuzzy modal variability can be expressed as

$$J_t(x,y)| = 2 \times \sum_{i=1}^{t}\sum_{j=1}^{m}\mu_{ij}^{r}(x,y)(1 - K(f_i(x,y), v_j(x,y))), \quad t \leq n. \quad (4)$$

The new membership function for any pixel at location (x,y) in $(t+1)^{th}$ frame can be computed as

$$\mu_{ij}^{t+1}(x,y) = \frac{(1/(1-K(f_i(x,y), v_j(x,y))))^{1/(r-1)}}{\sum\limits_{k=1}^{m}(1/(1-K(f_i(x,y), v_k(x,y))))^{1/(r-1)}}, \quad (5)$$

and the representative of each mode will be computed as

$$v_j^{t+1}(x,y) = \frac{\sum\limits_{i=1}^{n}\mu_{ij}^{r}(x,y)K(f_i(x,y), v_j(x,y))f_i(x,y)}{\sum\limits_{i=1}^{m}\mu_{ij}^{r}(x,y)K(f_i(x,y), v_j(x,y))}. \quad (6)$$

In the proposed scheme, the background model is initialized by considering a small region of support at a particular pixel location (x,y). The kernelized fuzzy modal variability is minimized initially with an assumption of the number of modes $(=2)$. The mode's representatives are computed by minimizing the kernelized fuzzy modal variation cost function. In the next stage, $f_{(t+1)}(x,y)$ is used to compute the kernelized fuzzy modal variation in two different ways for pixel location (x,y) in a new frame at $(t+1)^{th}$ frame. One assuming that $f_{(t+1)}(x,y)$ belongs to the existing background types and the other one assuming that $f_{(t+1)}(x,y)$ is a new background type. Hence, for the first kind, the new cost value using the fuzzy modal variation is computed by assuming that the new pixel belongs to each cluster/mode with fuzzy membership value. Hence, for considering the belongingness of the pixel to existing m background types a new cost value is computed. A separate cost value is also computed by considering the new pixel belonging to a new background type. Then, the minimum of the computed two cost values (one due to existing m background types and the other one with the newly created background types using $m+1$) is obtained and the old cost value is updated with the new one. Hence the representative corresponding to new modes is computed as

$$v_j(x,y)^{t+1} = \begin{cases} \frac{\sum\limits_{i=1}^{t}\mu_{ij}^{r}(x,y)(K(f_i(x,y), v_j(x,y)))f_i(x,y)}{\sum\limits_{i=1}^{m}\mu_{ij}^{r}(x,y)K(f_i(x,y), v_j(x,y))}, \\ \quad \text{if } f_i(x,y) \ \in \text{old background}; \\ \text{initiate } v_{m+1}(x,y), \\ \text{otherwise} \end{cases}. \quad (7)$$

where the new modes are computed as

$$v_{m+1}(x,y)^{t+1} = \eta_{med}(x,y), \quad (8)$$

η_{med} represents the median of pixel values considered in a small neighborhood of $f(x, y)$. The membership values are recomputed as

$$\mu_{ij}(x,y)^{t+1} = \begin{cases} \frac{(1/(1-K(f_i(x,y),v_j(x,y))))^{1/(r-1)}}{\sum_{k=1}^{m}(1/(1-K(f_i(x,y),v_k(x,y))))^{1/(r-1)}}, \\ \text{if } f_i(x,y) \ \in \text{old background types;} \\ \frac{(1/(1-K(f_i(x,y),v_j(x,y))))^{1/(r-1)}}{\sum_{k=1}^{m+1}(1/(1-K(f_i(x,y),v_k(x,y))))^{1/(r-1)},} \\ \text{otherwise} \end{cases}. \quad (9)$$

and the new cost function is calculated as

$$J_{t+1}(x,y) = \begin{cases} 2\sum_{i=1}^{t+1}\sum_{j=1}^{m}\mu_{ij}^{r}(x,y)(1-K(f_i(x,y),v_j(x,y))), \\ \text{if } f_i(x,y) \in \text{old background types;} \\ 2\sum_{i=1}^{t+1}\sum_{j=1}^{m+1}\mu_{ij}^{r}(x,y)(1-K(f_i(x,y),v_j(x,y))), \\ \text{otherwise.} \end{cases}. \quad (10)$$

This process is repeated over all the background training frames to construct the background model. In the next stage the object separation phase is followed.

B. Object Separation and Background Update

A similar procedure for separation of the foreground from the background in the target frame like background construction is followed. For each selected target frame, the belongingness of a particular pixel to each mode of background type constructed in the previous stage is tested using the cost function of the kernelized fuzzy modal variation. For each pixel location, the pixel value $f_t(x, y)$, for any frame at time instant, $t > n$ are fitted into the cost function as expressed in eq (1). For fitting the pixel value at each location $f_t(x, y)$, it is tested if it belongs to any of the previously constructed background type v_j, and then detected as the background pixel; else it is detected as a part of the moving object. It can be described in mathematical form as

$$\mathrm{O}_i(x,y) = \begin{cases} 0, \ \text{if} \\ \quad f_i(x,y) \in V_j, \ i = n+1, n+2, ..., N; \\ 1, \\ \quad \text{otherwise.} \end{cases} \quad (11)$$

Here, O_i is considered as the same size as that of the image frame. In the next stage of the proposed scheme, the background is updated.

There are two background model updating mechanisms reported in the literature of background subtraction; namely, conservative and blind [22]. In conservative updating mechanism, a pixel belonging to the foreground region in the target frame is never used for updating the background model; and the pixel belonging to the background is updated in each target frame after foreground detection. Similarly in blind updating mechanism, a pixel identified either as a part of background or foreground will be used for updating the background [22]. In most of the literature, the conservative updating mechanism is popularly used and hence, in the proposed scheme it is used. For each location (x, y), if $O_i = 0$, the corresponding pixel $f_i(x, y)$ is considered as background class and background model is updated as:

$$v_j(x,y) = \begin{cases} \frac{\sum_{i=1}^{t}\mu_{ij}^{r}(x,y)K(f_i(x,y),v_j(x,y))f_i(x,y)}{\sum_{i=1}^{m}\mu_{ij}^{r}(x,y)K(f_i(x,y),v_j(x,y))}, \\ \quad \text{if } f_i(x,y) \in v_j \quad j = 1,2,....,m; \\ v_j(x,y), j = 1,2,.....,m+1 \\ \text{otherwise;} \end{cases}. \quad (12)$$

and the membership values are updated as

$$\mu_{ij}(x,y) = \begin{cases} \frac{(1/(1-K(f_i(x,y),v_j(x,y))))^{1/(r-1)}}{\sum_{k=1}^{m}(1/(1-K(f_i(x,y),v_k(x,y))))^{1/(r-1)}} \\ \text{if } f_i(x,y) \in v_j \quad j = 1,2,....,m; \\ \mu_{ij}, j = 1,2,....,m+1, \\ \text{otherwise;} \end{cases}. \quad (13)$$

and the new cost function is calculated as

$$J(x,y) = \begin{cases} 2\sum_{i=1}^{t+1}\sum_{j=1}^{m}\mu_{ij}^{r}(x,y)(1-K(f_i(x,y),v_j(x,y))), \\ \text{if } f_i(x,y) \in v_j \quad j = 1,2,....,m; \\ J(x,y), j = 1,2,....,m+1, \\ \text{otherwise.} \end{cases} \quad (14)$$

IV. Experimental Results and Discussion

The proposed scheme is run on a *Core* i3^{TM} system with 4GB RAM, 3MB L2 cache. It is implemented by C++ programming language with Ubuntu operating system. The proposed scheme is tested on several sequences and for page constraint the visual results are shown on few sequences and quantitative results are shown on all the videos of two major challenging video databases: *Changedetection.net* and *Star*. Further, this section is divided into three parts: visual analysis of results, quantitative evaluation and discussions and future work.

A. Visual Results Analysis

In this section, the effectiveness of the proposed scheme is demonstrated on seven challenging sequences collected from different databases. These sequences are: *Water surface*, *MSA*, *Waving tree*, *PETS2006*, *Fountain02*, *Overpass* and *Badminton*. The considered sequences include different challenging effects: non-static background, vegetation changes, camera jitter, haze, noise, blurred, etc.

All the results reported on the above said sequences are provided in Fig. 2 (a). The results obtained by the GMM based BGS scheme shows the moving object region with many missing parts and are shown in Fig. 2(b). Mostly GMM based scheme is unable to detect the moving objects in the scene with non-static background. The KDE based scheme provided results as shown in Fig. 2(c). It is observed that on the non-static background sequences many missed alarms are obtained and for low resolution sequences false alarms are obtained.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

Fig. 2(d) shows the results of BRPCA scheme, where the proper object locations are not obtained. The results obtained by ViBe scheme are provided in Fig. 2(e), where a better object location is obtained in most of the sequences; except for the sequences with non-static background. The results obtained by pROST scheme as shown in Fig. 2(f) provide better results with less missed alarms. The results obtained by DPGMM and feature bag based BGS schemes shown in Figs. 2(g)-(h) have less missed alarms. However, the results obtained by the proposed BGS scheme as shown in Fig. 2(i) is found to have better accuracy with less misclassification error.

B. Quantitative Evaluation

One of the common ways of evaluating the results of any BGS scheme is by checking its effectiveness through quantitative evaluation measures. Quantitative evaluation measures evaluate the BGS scheme in an objective way. To evaluate the proposed scheme in a quantitative way, three evaluation measures: precision, recall and F-measure [46] are used. For calculating precision, recall and F-measure, pixel by pixel comparison of the ground-truth image is made with the output. In the experiment, we have considered the average of precision, average of recall and average of F-measure over the entire frames to evaluate the proposed scheme. All these measures are named as average precision, average recall, and average F-measure. It may be noted that higher the value of average precision, average recall and average F-measure, better is the results.

Average precision, average recall, and average F-measure obtained for all the considered sequences are provided in Table I. It may be observed from the table that higher values of average precision, average recall, and average F-measure are obtained for most of the sequences by the proposed BGS technique as compared to the other existing state-of-the-art BGS techniques.

In Table I, the threshold value considered for the manual thresholding based BGS scheme for the entire sequence are as follows: Water surface $Th = 40$, MSA $Th = 38$, Waving tree $Th = 30$, PETS2006 $Th = 50$, Fountain02 $Th = 35$, Overpass $Th = 40$, and Badminton $Th = 25$.

C. Evaluation on Large Databases

The effectiveness of the proposed scheme is tested on different large databases. In this article, two benchmark databases: changedetection.net[1] and Star [47] are used. The performance of the proposed scheme on these considered databases are provided as follows.

1) Change detection.net: It is one of the popular benchmark databases to test the effectiveness of any BGS. The changedetection.net database contains 31 sequences with different challenging conditions which include: baseline, dynamic background, camera jitter, intermittent object motion, shadow and thermal. This database contains image frames with the corresponding ground-truths. The performance of the proposed scheme on all the considered sequences are provided in Table II. The average of F-measures computed from all the sequences and all the categories are reported in Table II.

[1]http://www.changedetection.net/

TABLE II
AVERAGE F-MEASURE FOR CHANGE DETECTION.NET

Techniques	Baseline	DyBG	Cam. Jitter	IOM	Shadow	Thermal
GMM [6]	0.825	0.633	0.5970	0.520	0.737	0.662
DPGMM [10]	0.929	0.814	0.748	0.542	0.813	0.813
KDE [13]	0.909	0.596	0.572	0.409	0.803	0.742
ViBe [22]	0.870	0.565	0.600	0.507	0.803	0.665
RPCA [31]	0.677	0.684	0.547	0.672	0.729	0.565
SuBSENSE [26]	0.950	0.818	0.815	0.657	0.899	0.817
SOBS-CF [39]	0.873	0.309	0.745	0.534	0.664	0.873
feature bags [17]	0.943	0.837	0.818	0.643	0.820	0.822
Multi-background [27]	0.932	0.621	0.836	**0.823**	0.838	**0.910**
WeSamBE [28]	0.936	0.790	0.780	0.724	0.914	0.813
DeepBS [34]	0.965	0.844	0.896	0.689	0.943	0.650
Cascade CNN [35]	0.967	**0.947**	**0.967**	0.868	0.946	0.887
Proposed	**0.968**	0.861	0.873	0.729	**0.947**	**0.901**

For evaluating the proposed scheme on this database, it is compared with ten existing state-of-the art techniques, which include GMM [6], DPGMM [10], KDE [13], ViBe [22], SuBSENSE [26], SOBS-CF [39], feature bags [17], multimode background [27], WeSamBE [28], DeepBS [34] and cascade CNN [35] schemes. It may be observed from this table that a better value of F-measure is obtained by the proposed scheme as compared to the ten state-of-the-art techniques. For sequences under the categories of baseline, dynamic background and shadow the average F-measure obtained by the proposed scheme is the highest as compared to all other sequences. However, for the camera jitter, inter object motion, and thermal images, the proposed scheme is found to provide comparable results as compared to the DeepBS and multimode background techniques. The proposed scheme is found to be providing a higher average F-measure value for all the categories as compared to the fuzzy based BGS SOBS-CF technique.

2) Star Database: Star database [47] is another popular database which is used to check the effectiveness of the proposed scheme. This database contains 8 challenging sequences which are affected by systematic noise and camera jitter. The proposed scheme is evaluated on this database and its effectiveness is verified by comparing it with five state-of-the-art BGS techniques. The BGS techniques used for evaluation are GMM [6], Li *et al.* [47], self organizing [48], statistical feature bag [17], and DPGMM [10] techniques. The performance of the proposed BGS is evaluated on this database by considering the average similarity measures obtained over the different sequences on this database.

Table III shows the evaluation measures obtained on the Star database by the five considered state-of-the-art techniques and the proposed technique. It may be observed from Table III, that a better accuracy for change detection is obtained by the proposed scheme on star database as compared to the other five state-of-the-art techniques. The proposed scheme is found to provide competitive results with feature bags and DPGMM techniques, respectively on curtain and bootstrap sequences.

D. Discussions and Future Work

The use of fuzzy concept in the proposed scheme incorporates a soft decision for deciding the belongingness of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

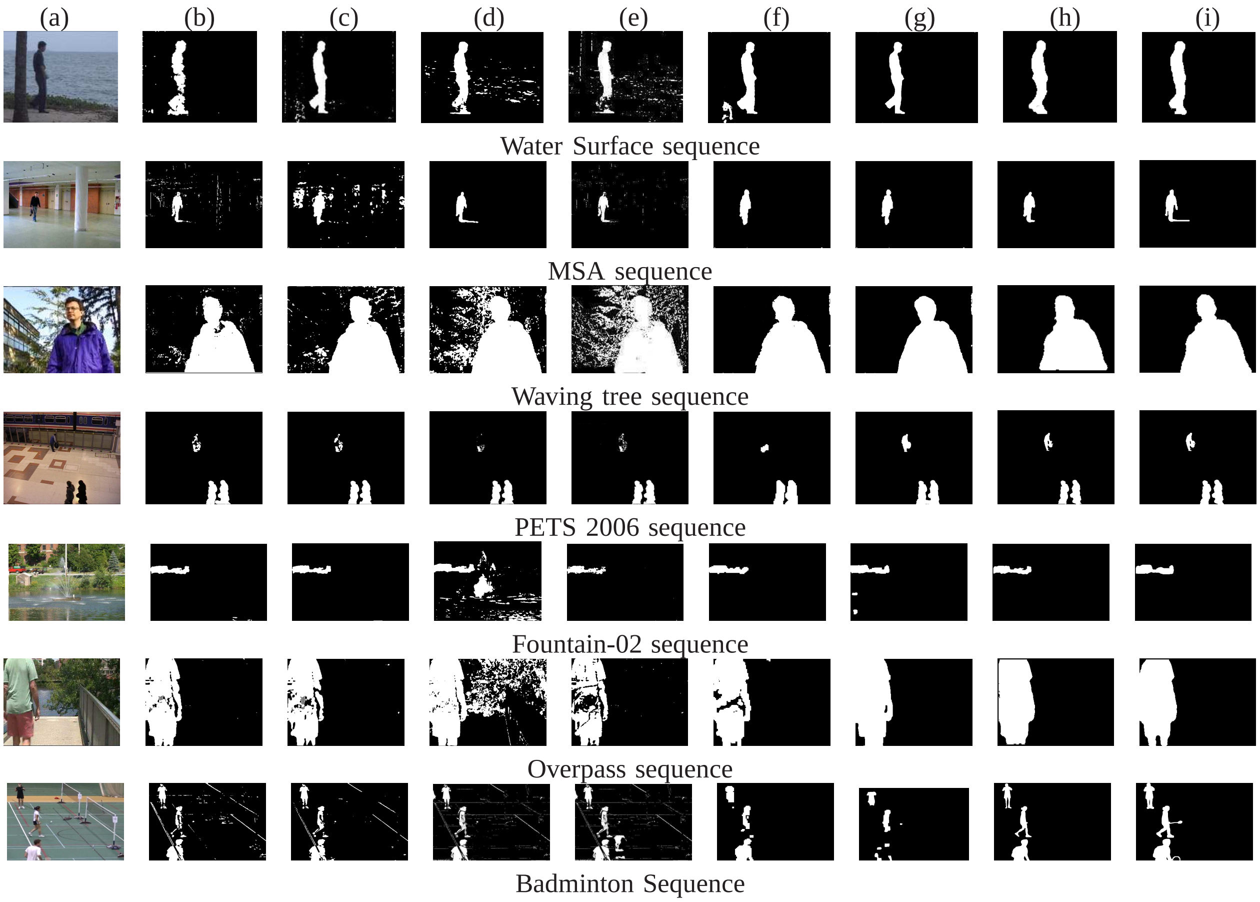

Fig. 2. Moving object detection for different sequences: (a) original frame, results obtained by BGS scheme based on (b) GMM, (c) KDE, (d) BRPCA, (e) ViBe, (f) pROST, (g) DPGMM, (h) feature bags and (i) proposed

TABLE III
AVERAGE SIMILARITY MEASURE FOR STAR DATABASE

	GMM [6]	Video plane [47]	SOM [48]	DPGMM [10]	feature bags [17]	Proposed
Campus	0.0757	0.1596	0.6960	0.7876	0.8011	**0.8112**
Fountain	0.6854	0.0999	0.6554	0.7424	0.7672	**0.7712**
Curtain	0.7580	0.1841	0.8178	0.8411	**0.8963**	0.8761
Lobby	0.6519	0.1554	0.6489	0.6665	0.8865	**0.9112**
Station	0.5363	0.5209	0.6677	0.6733	0.6674	**0.6781**
Airport	0.3335	0.1135	0.5943	0.5676	0.6011	**0.6202**
Bootstrap	0.3838	0.3079	0.6019	**0.6496**	0.6238	0.6381
Escalator	0.1388	0.1294	0.5770	0.5522	0.5610	**0.5772**

TABLE IV
AVERAGE F-MEASURE OBTAINED BY PROPOSED SCHEME FOR CHANGE DETECTION.NET DATA WITH DIFFERENT %AGE OF TRAINING

Training	Baseline	DyBG	Camera Jitter	IOM	Shadow	Thermal
2%	0.915	0.852	0.801	0.699	0.893	0.865
3%	0.934	0.857	0.834	0.711	0.912	0.878
5%	0.968	0.861	0.873	0.729	0.947	0.901
7%	0.973	0.873	0.881	0.732	0.949	0.912
10%	0.981	0.875	0.881	0.733	0.949	0.921

pixels to foreground/background. The use of fuzzy kernel in the proposed scheme, helps in detecting the local changes in a scene containing blurred boundary, light shadow, low illuminated environmental condition, noisy scene, objects with dynamic background, etc. In this context, fuzzy set theories are reputed to handle uncertainties to a reasonable extent, arising from deficiencies of information available from a situation (the deficiency may result from incomplete, ill-defined, not fully reliable, vague and contradictory information).

All the results reported in Fig. 2 and Tables I- III are obtained by considering a total 5% of the training data. In order to test the efficacy of the proposed scheme, it is tested it with different percentage of training samples and obtained that the proposed scheme provides an acceptable accuracy with 5% training samples. The results obtained with different percentages of training samples are reported in Table IV. The average execution time per frame for training and testing of each frame of the considered seven sequences are presented in Table V.

One of the important parameters considered in this experiment is r, the fuzzifier. All the results reported in this manuscript are obtained by the optimal value (!) of r (set by trial and error basis). Higher the value of this parameter, greater is the amount of fuzziness to the cost function of the fuzzy modal variation. Similarly, a lesser value of it will provide lesser fuzziness to the cost function of the fuzzy modal variation. Although r can be defined taking into account the changes in a scene in real life scenarios, its choice can be critical. A smaller value of this parameter may produce fragments in foreground/background regions due to decrease

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

TABLE I
AVERAGE PRECISION, RECALL AND F-MEASURE ON ALL FRAMES FOR DIFFERENT IMAGE SEQUENCES

	Water surface			MSA			Waving tree			PETS2006			Fountain02			Overpass			Badminton		
	No. of Frames: 582			No. of Frames: 582			No. of Frames: 1045			No. of Frames: 1200			No. of Frames: 1499			No. of Frames: 3000			No. of Frames: 1150		
Approaches	Pr	Re	FM	Pr	Re	FM	Pr	Re	FM	Pr	Re	FM	Pr	Re	FM	Pr	Re	FM	Pr	Re	FM
Manual threshold	0.46	0.72	0.55	0.8	0.78	0.79	0.71	0.80	0.76	0.66	0.93	0.77	0.44	0.85	0.58	0.61	0.76	0.68	0.58	0.67	0.62
GMM [6]	0.86	0.85	0.85	0.80	0.70	0.75	0.70	0.80	0.75	0.78	0.87	0.83	0.75	0.87	0.80	0.92	0.83	0.87	0.63	0.75	0.69
Radiometric similarity [15]	0.88	0.84	0.86	0.81	0.84	0.83	0.76	0.83	0.80	0.68	0.90	0.77	0.47	0.75	0.58	0.71	0.75	0.73	0.59	0.71	0.65
Wronskian [5]	0.68	0.85	0.75	0.80	0.83	0.81	0.73	0.84	0.78	0.72	0.98	0.83	0.43	0.86	0.57	0.62	0.79	0.69	0.59	0.74	0.68
Codebook [14]	0.51	0.71	0.60	0.81	0.86	0.84	0.80	0.88	0.84	0.84	0.97	0.90	0.52	0.85	0.64	0.78	0.78	0.78	0.63	0.80	0.71
KDE [13]	0.40	0.79	0.52	0.67	0.79	0.73	0.53	0.81	0.64	0.83	0.79	0.81	0.85	0.80	0.82	0.80	0.85	0.82	0.67	0.79	0.73
BRPCA [19]	0.86	0.92	0.89	0.83	0.90	0.86	0.81	0.86	0.83	0.80	0.86	0.83	0.52	0.86	0.65	0.67	0.81	0.73	0.69	0.75	0.72
DT [9]	0.84	0.91	0.88	0.80	0.88	0.84	0.80	0.75	0.77	0.77	0.81	0.79	0.81	0.82	0.76	0.74	0.78	0.76	0.67	0.78	0.72
ViBe [22]	0.71	0.85	0.77	0.86	0.90	0.88	0.79	0.84	0.82	0.86	0.70	0.78	0.86	0.80	0.83	0.85	0.76	0.80	0.71	0.79	0.75
Gaussian Wronskian [3]	0.91	0.95	0.93	0.83	0.87	0.85	0.77	0.84	0.81	0.90	0.97	0.93	0.80	0.79	0.79	0.67	0.79	0.72	0.67	0.71	0.69
pROST [21]	0.65	0.78	0.72	0.85	0.93	0.89	0.79	0.88	0.83	0.70	0.67	0.68	0.80	0.89	0.84	0.90	0.88	0.89	0.87	0.81	0.84
DPGMM [10]	0.89	0.94	0.92	0.87	**0.94**	0.90	0.82	0.87	0.85	0.85	**0.98**	0.91	0.86	**0.96**	0.90	0.92	**0.99**	**0.94**	0.88	0.68	0.78
feature bags [17]	0.92	**0.97**	0.95	0.90	0.92	0.91	0.86	**0.96**	0.89	0.92	**0.98**	0.95	0.89	0.90	0.89	0.92	0.93	0.92	0.81	0.84	0.85
Proposed	**0.94**	**0.97**	**0.95**	**0.93**	0.92	**0.92**	**0.88**	0.92	**0.90**	**0.94**	0.96	**0.95**	**0.92**	0.93	**0.92**	**0.93**	0.93	0.92	**0.88**	**0.82**	**0.85**

Pr:precision, Re:recall and FM:F-measure.

TABLE V
AVERAGE EXECUTION TIME (IN SECOND) PER FRAME REQUIRED FOR TRAINING AND TESTING OF THE PROPOSED SCHEME

Water surface		MSA		Waving tree		PETS2006		Fountain02		Overpass		Badminton	
No. of Frames: 528		No. of Frames: 582		No. of Frames: 1045		No. of Frames: 1200		No. of Frames: 1499		No. of Frames: 3000		No. of Frames: 1150	
Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
0.05	0.07	0.07	0.09	0.07	0.09	0.09	0.11	0.08	0.10	0.07	0.09	0.09	0.11

in fuzziness. Larger value of it may produce de-fragmented smoothed regions in foreground. In the considered experiment, the value of r is assumed to be in the range $[2, 5]$. The results on all the test sequences considered in this experiment are reported with $r = 3$. The other parameter considered in the experiment is window size w. The size of the window impacts in smoothing the boundary of the changed regions in the foreground/background map. It is observed that as the value of w increases, the silhouette of the object increases. It is also observed that w size with 5×5 provides better results for most of the considered sequences.

There is a probability that after detecting the local changes in a video scene using the proposed methodology, some parts of the unchanged regions can be identified as part of the changed regions and vice-versa. These are found to be very small isolated points and arise due to noise or vegetation changes. In order to remove such background blobs, a connected-component labeling [49] technique is employed.

Although the proposed scheme yields better results for the considered test sequences presented, however, at higher dynamism, very low contrast and dense haze environment condition, it does not perform well. In underwater scenarios, the proposed scheme also suffers from decolorization of the scene of view. In future work, these challenges will be focussed. To deal with such underwater vagueness, some approaches using fuzzy set based dehazing schemes fused with fuzzy histogram mechanism will be used for background modeling.

V. CONCLUSIONS

In this article, a novel background subtraction scheme is proposed, where fuzzy set theoretic decision process is utilized for detecting the local changes. In the proposed scheme, a kernelized fuzzy modal variation mechanism is used to detect the local changes from a sequence of image frames. The proposed technique is tested on several image sequences and for page constraint, the results on two databases are reported including seven test image sequences. The proposed scheme is found to be robust against noise, illumination variation, surface reflection, and uniform shading. It is found to provide better results against non-static background and also can handle camera jitter. The proposed scheme is tested against different sequences with several challenges including two popular databases: changedetcetion.net and star. The proposed scheme is compared with twenty one state-of-the-art background subtraction schemes and found to be efficient. To corroborate the findings, three performance evaluation measures: precision, recall and F-measure are used. It is observed that the proposed scheme provides higher values of these measures for most of the test cases.

REFERENCES

[1] A. M. Tekalp, *Digital Video Processing*. New Jersey: Prentice Hall, 1995.
[2] B. N. Subudhi, P. K. Nanda, and A. Ghosh, "A change information based fast algorithm for video object detection and tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 993–1004, 2011.
[3] B. N. Subudhi, S. Ghosh, and A. Ghosh, "Change detection for moving object segmentation with robust background construction under Wronskian framework," *Machine Vision and Applications*, vol. 24, no. 4, pp. 795–809, 2013.
[4] R. F. Gonzalez and R. E. Woods, *Digital Image Processing*. Singapore: Pearson Education, 2001.
[5] E. Durucan and T. Ebrahimi, "Change detection and background extraction by linear algebra," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1368–1381, 2001.
[6] C. Stauffer and W. Grimson, "Learning patterns of activity using real time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–767, 2000.
[7] M. Chen, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal background subtraction using minimum spanning tree and optical flow," in *13th European Conference on Computer Vision*, 2014, pp. 521–534.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2938342, IEEE Transactions on Multimedia

[8] D. K. Rout, B. N. Subudhi, T. Veerakumar, and S. Chaudhury, "Spatio-contextual Gaussian mixture model for local change detection in underwater video," *Expert Systems with Applications*, vol. 97, pp. 117 – 136, 2018.

[9] A. Chan, V. Mahadevan, and N. Vasconcelos, "Generalized Stauffer and Grimson background subtraction for dynamic scenes," *Machine Vision and Applications*, vol. 22, pp. 751–766, 2011.

[10] T. S. F. Haines and T. Xiang, "Background subtraction with Dirichlet process mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 670–683, 2014.

[11] P.-M. Jodoin, M. Mignotte, and J. Konrad, "Statistical background subtraction using spatial cues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1758–1763, 2007.

[12] X. Zhang, C. Zhu, H. Wu, Z. Liu, and Y. Xu, "An imbalance compensation framework for background subtraction," *IEEE Transactions on Multimedia*, vol. 19, no. 11, pp. 2425–2438, 2017.

[13] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[14] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-time Imaging*, vol. 11, no. 3, pp. 167–256, 2005.

[15] P. Spagnolo, T. D. Orazio, M. Leo, and A. Distante, "Moving object segmentation by background subtraction and temporal analysis," *Image and Vision Computing*, vol. 24, no. 5, pp. 411–423, 2006.

[16] N. A. Mandellos, I. Keramitsoglou, and C. T. Kiranoudis, "A background subtraction algorithm for detecting and tracking vehicles," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1619 – 1631, 2011.

[17] B. N. Subudhi, S. Ghosh, S. C. Shiu, and A. Ghosh, "Statistical feature bag based background subtraction for local change detection," *Information Sciences*, vol. 366, pp. 31 – 47, 2016.

[18] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*,, vol. 28, no. 4, pp. 657 –662, 2006.

[19] E. Z. C. Guyon, T. Bouwmans, "Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis," *INTECH*, *Principal Component Analysis*, vol. 1, pp. 223–238, 2012.

[20] J. Seo and S. D. Kim, "Recursive on-line $(2\mathrm{D})^2 PCA$ and its application to long-term background subtraction," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2333–2344, 2014.

[21] F. Seidel, C. Hage, and M. Kleinsteuber, "pROST - a smoothed lp-norm robust online subspace tracking method for realtime background subtraction in video," *Machine Vision and Applications*, vol. 25, no. 5, pp. 1227–1240, 2014.

[22] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[23] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2502–2514, 2015.

[24] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp. 1726–1740, 2018.

[25] S. Minaee and Y. Wang, "An ADMM approach to masked signal decomposition using subspace representation," *IEEE Transactions on Image Processing*, 2019.

[26] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "SuBSENSE: a universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, 2015.

[27] H. Sajid and S. C. S. Cheung, "Universal multimode background subtraction," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3249–3260, 2017.

[28] S. Jiang and X. Lu, "WeSamBE: a weight-sample-based method for background subtraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2105–2115, 2017.

[29] H. Lee, H. Kim, and J. Kim, "Background subtraction using background sets with image and color space reduction," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2093–2103, 2016.

[30] X. Liu, J. Yao, X. Hong, X. Huang, Z. Zhou, C. Qi, and G. Zhao, "Background subtraction using spatio-temporal group sparsity recovery," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 8, pp. 1737–1751, 2018.

[31] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4075–4090, 2016.

[32] M. Chen, X. Wei, Q. Yang, Q. Li, G. Wang, and M. Yang, "Spatiotemporal GMM for background subtraction with superpixel hierarchy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1518–1525, 2018.

[33] M. D. Gregorio and M. Giordano, "Background estimation by weightless neural networks," *Pattern Recognition Letters*, vol. 96, pp. 55 – 65, 2017.

[34] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognition*, vol. 76, pp. 635 – 649, 2018.

[35] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognition Letters*, vol. 96, pp. 66–75, 2017.

[36] M. H. Sigari, N. Mozayani, and H. R. Pourreza, "Fuzzy running average and fuzzy background subtraction: Concepts and application," *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 138–143, 2008.

[37] H. Zhang and D. Xu, "Fusing color and texture features for background model," in *Fuzzy Systems and Knowledge Discovery*, L. Wang, L. Jiao, G. Shi, X. Li, and J. Liu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 887–893.

[38] F. El-Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," in *IEEE International Conference on Fuzzy Systems*, 2008, pp. 1729–1736.

[39] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Computing and Applications*, vol. 19, no. 2, pp. 179–186, 2010.

[40] W. Kim and C. Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms," *IEEE Signal Processing Letters*, vol. 19, no. 3, pp. 127–130, 2012.

[41] M. Balcilar and A. C. Sonmez, "Region based fuzzy background subtraction using choquet integral," in *Adaptive and Natural Computing Algorithms*, M. Tomassini, A. Antonioni, F. Daolio, and P. Buesser, Eds., 2013.

[42] T. Bouwmans and F. El-Baf, "Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models," *MASAUM Journal of Basic and Applied Sciences*, vol. 1, no. 2, pp. 265–277, 2009.

[43] A. Mondal, B. N. Subudhi, M. Roy, S. Ghosh, and A. Ghosh, "A study on nonlinear classifier-based moving object tracking," in *Intelligent Computing, Communication and Devices*, L. C. Jain, S. Patnaik, and N. Ichalkaranje, Eds. New Delhi: Springer India, 2015, pp. 571–578.

[44] A. Ghosh, B. N. Subudhi, and S. Ghosh, "Object detection from videos captured by moving camera by fuzzy edge incorporated Markov random field and local histogram matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1127–1135, 2012.

[45] D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," *Fuzzy Sets and Systems*, vol. 161, pp. 522–543, 2010.

[46] B. N. Subudhi, S. Ghosh, and A. Ghosh, "Application of Gibbs Markov random field and Hopfield-type neural networks for detecting moving objects from video sequences captured by static camera," *Soft Computing*, vol. 19, pp. 2769–2781, 2015.

[47] J. Li and X. Zhang, "Video object plane extraction for surveillance application," in *Proceedings of the IEEE Third International Conference on Machine Leaning and Cybernatics*, 2004, pp. 3928–3931.

[48] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, 2008.

[49] D. H. Parks and S. S. Fels, "Evaluation of background subtraction algorithms with post-processing," in *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, 2008, pp. 192–199.