Salient Features for Moving Object Detection in Adverse Weather Conditions during Night Time

Anu Singha and Mrinal Kanti Bhowmik*

Abstract—**Foreground segmentation of moving objects in adverse atmospheric conditions such as fog, rain, low light and dust is a challenging task in computer vision. The advantages of thermal infrared imaging at night time under adverse atmospheric conditions have been demonstrated, which are due to the long wavelength. However, existing state-of-the-art object detection techniques have not been useful in such scenarios. In this paper, we propose an improved background model that utilizes both thermal pixel intensity features and spatial video salient features. The proposed spatial video salient features are represented as an Akin-based per-pixel Boolean string over a local region block, and depend on the effect of neighbouring pixels on a centre pixel. The result of this Boolean procedure is referred to as the - '*Akin-based Local Whitening Boolean Pattern (ALWBP),*' which differentiates foreground and background region accurately, even against a cluttered background. The background model is controlled via (i) the automatic adaptation of parameters such as the decision threshold R_T and, learning parameter L, and (ii) the updating of background samples B_{sample_int} and,- B_{sample_ALWBP} to minimize (a) the effect of the background dynamics of outdoor scenes, and (b) the temperature polarity changes during the maiden appearance of a moving object in thermal frame sequences. The performance of this model is evaluated using nine existing standard segmentation performance metrics on our newly created -'*Tripura University Video Dataset at Night time (TU-VDN)*' and on the publicly available CDnet-2014 dataset. Our newly created weather-degraded video dataset, namely, TU-VDN, consists of sixty video sequences that represent four atmospheric conditions, namely, low light, dust, rain, and fog. The results of a performance comparison with fourteen state-of-the-art detection techniques also demonstrate the high accuracy of the proposed technique.**

***Index Terms*— Atmosphere, aerosols, infrared, Akin, whitening, Boolean pattern, background model.**

I. INTRODUCTION

Automatic night vision systems for the intelligent monitoring of moving objects assume that the input images have clear visibility under lane light; however, unfortunately, this assumption does not always hold [1]. The moving object monitoring performance depends closely on the enhanced quality of the images [2]. The quality of outdoor images is affected by several atmospheric conditions that alter the key characteristics (e.g., intensity, colour, polarization, and coherence) of the light source due to scattering by medium aerosols [3, 4]. Although computer vision systems perform well in indoor or outdoor environments during the day time, they encounter issues in outdoor atmosphere-affected environments. A satisfactory solution for night time is highly necessary because darkness causes major safety problems due to the collision of objects [5, 6]. The poor appearance of night images under subjective lighting and atmospheric conditions is a general problem for analysis in computer vision [7]. Due to adverse atmospheric conditions, the contrast of the images is degraded, which affects the visibility in such a scenario. The contrast degradation depends on the coefficient of light scattering through aerosols that are suspended in the atmosphere. In the last few decades, large datasets have been designed to meet the increasing demands for the development of new models for object detection under poor atmospheric conditions [8, 9]. However, there is still a lack of video datasets for moving object detection tasks that provide balanced coverage in atmosphere-degraded outdoor scenes, especially at night.

Furthermore, for detecting moving objects, both a visual digital camera and a typical charge-coupled device (CCD) camera have the advantage of high resolution, which renders them more suitable for day time or night time use with a proper lighting setup. However, they are ineffective in environments with poor illumination or visibility due to atmospheric conditions because the appearance of objects in the captured images is not as clear as in images that are captured during under normal atmospheric conditions [1, 10]. Several related works have been conducted in such environments [11, 12, 13]. To address the limitations of visual and CCD cameras at night time, many studies have been conducted on methods that detect objects with near/far-infrared (NIR/FIR) based cameras [14, 15, 16]. NIR cameras are robust against darkness, and however, they have a similar drawback to that faced by CCD cameras when the interferences are produced by vehicle headlights. In addition, the attenuation of visual, CCD, and NIR radiation that is produced through atmospheric aerosols is mostly due to their short wavelengths. In contrast, FIR cameras enable robust object detection regardless of the atmospheric conditions because as the spectrum wavelength increases, the effect of bad atmospheric conditions decreases [4]. However, there have many key issues that are related to object detection at night using an FIR camera, such as the following: (i) **Flat Cluttered Background:** The infrared radiation signal must travel from the target to the camera sensor among adverse atmospheric particles and is attenuated due to scattering; the loss of radiation along the way produces a blurred flat region. In addition, with the thermal sensors, because of large variations in the surface, which includes hot and cool objects such as buildings, vehicles, animals, humans, and light poles, the foreground objects and the background scene become indistinguishable; (ii) **Temperature Polarity Changes:** Thermal temperature adjustment during the maiden appearance of a moving object in a video sequence causes illumination-type effects in the background model from the current video frame and, therefore, yields false classifications. (iii) **Background Dynamics:** Outdoor scenes are affected by

This paper was submitted on March 26th, 2019 and accepted on June 24th, 2019.

Anu Singha, and Mrinal. K. Bhowmik, are with the Department of Computer Science and Engineering, Tripura University, Suryamaninagar, Tripura, 7990022, India (e-mail: anusingh5012@gmail.com; mrinalkantibhowmik@tripurauniv.in/ mkb.cse@gmail.com)

ACKNOWLEDGEMENT: The work that is presented here is being conducted in the Computer Vision Laboratory of Computer Science and Engineering Department of Tripura University (A Central University), Tripura, Suryamaninagar-799022.

* asterisk represent the corresponding author.

movement in the background, e.g., due to waves or swaying tree leaves.

The simplest object detection strategy is to segment moving regions of interest from the static background. The traditional background/foreground segmentation methods may be ill-suited for overcoming outdoor environment issues such as dynamic behaviour in background (e.g., swaying trees) due to the key issues that are discussed above. The background-model-based background segmentation methods are mostly pixel-level approaches that are either parametric [17] or non-parametric [18, 19, 20]. In FIR imaging, the pixel-intensity-based methods are not well suited for differentiating flat regions (with similar intensities between the foreground and background). It is necessary to use spatial features to analyse the texture [21, 22], especially when cluttered backgrounds are involved.

The overall workflow of this paper is as follows. First, we briefly describe the night dataset that we created under various poor atmospheric conditions. The dataset was captured throughout the year using an FLIR thermal camera. The thermal video clips of outdoor night scenes are typically affected by bad atmospheric conditions, which result in blurred thermal sequences. To keep the complexity of the object-segmentation-based detection methods minimal, we used a deblurring pre-processing technique, namely, *blind deconvolution*, prior to segmentation [23]. In the background segmentation section, we propose an improved non-parametric background model that uses both local textures and thermal pixel intensities to discriminate between the foreground and background of flat cluttered regions. This novel non-parametric approach also handles incorrect classifications that are caused by dynamic background and temperature polarity changes. To regularize the salt-and-pepper noise segmentation results, we use the *Markov random field (MRF)* graphical model. The noisy scattered segment pixels will connect geometrically according to their closeness in the MRF graph [24]. Then, the captured dataset is evaluated via our proposed method, and eleven state-of-the-art approaches. According to the results, our method outperforms these state-of-the-art methods in terms of three performance metrics – accuracy, F_1-score, and Matthews correlation coefficient (MCC). The proposed method is also evaluated on the changeDetection.net (CDnet) 2014 dataset [9].

The primary contributions of this paper are summarized as follows: **(1)** The paper describes in brief a comprehensive thermal video dataset of outdoor night scenes that are degraded by various adverse weather conditions, such as fog, dust, rain, and low light/poor illumination. This dataset is referred to as *Tripura University Video Dataset at Night time (TU-VDN)*. Researchers can utilize this dataset for testing and ranking of existing and new algorithms for moving object detection; (*Dataset is available for the research community, contact email or website respectively: mrinalkantibhowmik@tripurauniv.in or mkb.cse@gmail.com and www.mkbhowmik.in.* **(2)** The paper proposes an improved video salient feature-based background model algorithm for detecting moving objects in night videos that were captured under adverse atmospheric conditions, in which thermal intensity information, in addition to spatial information, is fully taken into account. This algorithm can handle key challenging issues in thermal and outdoor adverse atmospheric environments, such as a flat cluttered background, a dynamic background, and thermal temperature polarity changes; **(3)** The proposed salient-feature-based moving object detection method is successfully applied to our adverse-atmospheric-condition-based thermal night dataset, namely, *TU-VDN*, and the results demonstrate that it outperforms related state-of-the-art methods in terms of detection performance; **(4)** The performance of the proposed method is also evaluated on change detection dataset - 'CDNet 2014'.

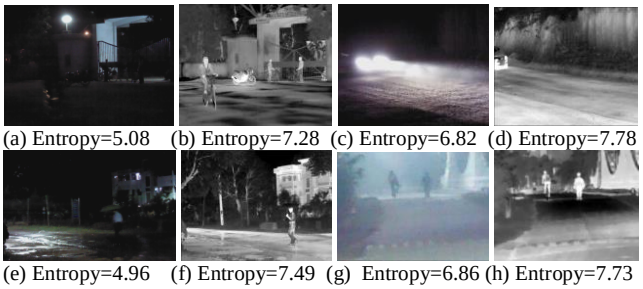
The remainder of this paper is organized as follows. In the next section, the dataset-capturing design, conditions, and statistics are described. The problem is defined in Section III, and the related literature is surveyed in Section IV. In Section V, an improved background segmentation algorithm that uses spatial features and thermal pixel intensities is presented. In Section VI, a complete evaluation of the captured dataset is presented, followed by a discussion of the experimental results of the proposed method and a performance comparison with state-of-the-art approaches. Finally, in Section VII, we present the conclusions of this work and discuss future work.

II. Brief Description of the TU-VDN Dataset

Atmospheric aerosols reduce the visibility of the targets in a scene. This effect is especially debilitating at night. It directly affects the visibility through the aerosols and through vehicle headlamps and, street headlamps. At night, an object is typically visible when light from a source is reflected by the object back to the terminal camera sensors. To detect the presence of objects, terminal sensors use several electromagnetic (EM) spectra that range from the visible to the near-infrared to the far-infrared regions. For electro-optical (EO) sensors, when an EM wave propagates through the atmosphere, the primary factors that are responsible for extinction are *absorption* and *scattering* by atmospheric aerosols (for example, -rain, dust, and fog). Both factors degrade the performance of all sensors [4]. Due to these poor atmospheric/weather conditions, the contrast of a scene is degraded, which affects the visibility. This degradation depends on the aerosols is as follows: as the aerosol size decreases, the amount of scattering increases. The relative amount of atmospheric-aerosol-based EM radiation attenuates according to the ratio of the droplet radius to the wavelength.

The atmosphere influences how far one can see through aerosols; the type of infrared camera that is used and the waveband in which the camera operates are also of importance. Because the particle size well exceeds the wavelength in the visible portion of the EM spectrum (0.4 to 0.74 μm), attenuation by atmospheric aerosols is independent of the wavelength. As the wavelength increases, attenuation becomes less of an issue. Since wavelengths in the far-infrared region exceed those of other infrared wave bands, impact of particles on far-infrared waves is relatively insignificant. Far-infrared waves provide the advantage of 'seeing' not only at night but also through many atmospheric aerosols such as dust, fog, and rain. Fig. 1 shows, visual frames and the corresponding thermal sample frames that were captured at night under several atmospheric conditions. To characterize the textures in night time visual and night time thermal images, we have used entropy to measure the contents, where a higher entropy value in night time thermal frames indicates an image with adequate details of information in terms 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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology



(a) Entropy=5.08 (b) Entropy=7.28 (c) Entropy=6.82 (d) Entropy=7.78

(e) Entropy=4.96 (f) Entropy=7.49 (g) Entropy=6.86 (h) Entropy=7.73

Fig. 1. Sample frames of the created dataset at night time (a), (b) a visual frame and the corresponding thermal frame, respectively, under low-light conditions; (c), (d) a visual frame and the corresponding thermal frame, respectively, under dust conditions; (e), (f) a visual frame and the corresponding thermal frame, respectively, under rain conditions; and (g), (h) a visual frame and the corresponding thermal frame, respectively, under fog conditions.

better quality. The level of distinguishable information of thermal frames is capable of revealing important hidden targets/objects than the night visual frames.

The most closely related datasets in the literature include thermal and visual-thermal frames since no dataset is available for purely night-based or poor atmospheric-condition-based scenarios such as dust, fog, and rain. Several of these datasets have been designed for evaluating moving object detection methods. Among these datasets, OSU-T [8], BU-TIV [25], ASL-TID [26], and LTIR [27] were captured using thermal sensors to detect and track objects, whereas the BU-TIV dataset is primarily designed for visual analysis tasks. These datasets only contain day-time video sequences, which have the challenges of cluttered background, occlusion, static and moving cameras, and object size variation, whereas the OSU-T dataset includes various weather conditions and was captured using a low-resolution thermal camera to detect only pedestrians. Numerous datasets, such as LITIV [28], AIC-TV [29], OSU-CT [30], CVC-14 [31], KAIST [32], CDNet 2012 [33], and CDNet 2014 [9], contain both colour and thermal video sequences; a few of them (LITIV, OSU-CT, and KAIST) fuse two modalities for robust detection. AIC-TV, CV-14, KAIST, and CDNet 2014 contain night video sequences. These datasets consist of various challenges, such as scale variations, lighting conditions, dynamic backgrounds, shadows, camera jitter, low frame rate, and turbulence. Very few datasets are considered adverse weather conditions; (i) OSU-T dataset is only considered for pedestrian detection in low resolution thermal imagery with only 10 number of sequences with total 284 images, (ii) CDNet 2014 is considered only day time weather conditions such as snowfall under vehicle and pedestrian detection. The video sequences contains of only four clips using visual camera. These datasets are contains of very limited video sequences of adverse weather conditions. Thus, it is difficult to evaluate the robustness of object detection methods under atmospheric conditions, especially for night vision, because more than half of object-related accidents occur at night. In contrast, our motive is to providing a new dataset comprising of several adverse weather conditions with large number of video sequences compared to existing datasets.

Therefore, we have designed a standard night-vision video dataset that is based on several atmospheric-weather-degraded conditions and covers many real-world scenarios. The considered atmospheric conditions are dust aerosols, fog aerosols, rain aerosols, and a low-light environment, under which we utilize a thermal camera. The dataset video recording conditions, dataset information, key features, and ground-truth annotation details are discussed in an article [34].

The TU-VDN dataset provides a realistic diverse set of outdoor videos in night vision that were captured via a thermal modality. The current dataset consists of 60 video sequences that were captured under various atmospheric conditions; the key challenges of the video clips are listed in Table II. Each video clip is 2 minutes in duration and was recorded with an FLIR camera that was rigidly mounted with 90^{o} alignments on a tripod stand by maintaining 200m to 2km distances from objects. In contrast, for a motion background, the video is captured by mounting the camera on a moving vehicle (20~30 km/h) such that the objects, camera, and background are moving simultaneously. Overall statistics has listed in Table I.

The key features of the designed dataset are as follows: **(i)** Each frame contains multiple types of moving objects, e.g., pedestrians, various types of vehicles, bicyclists, motorbikes, trains, and pets; **(ii)** The night video clips were captured under three outdoor atmospheric scenarios, namely, dust, rain, and fog, which produce flat regions in thermal scenes. In addition, the captured scenes are mostly in urban areas, which correspond to larger surface variations due to the presence of hot and cool objects such as houses, warehouses, office buildings, streets, and residents. Therefore, areas with varied background and adverse weather conditions produce thermal characteristics that lead to an increased *flat cluttered* region in the target area; **(iii)** A conventional challenge is encountered, namely, *a dynamic background* due to shaking trees, since the whole dataset was recorded in an outdoor environment; **(iv)** The key issue with the FIR camera is *thermal temperature adjustment* during the maiden appearance of a moving object in a video sequence, which causes illumination-type effects in the background model from the current video frame; **(v)** Motion-camera-based videos are captured by mounting the camera on a moving vehicle, where the camera and objects are moving and shaking simultaneously.

III. Problem Definition

The thermal infrared radiation signal must travel from the target to the camera detector sensor under adverse weather conditions or through atmospheric particles; therefore, more of

TABLE I: Statistics of created dataset in different atmospheric conditions at night time.

Image Type	Camera Model	Camera Situation	Background Condition	Atmospheric Conditions				Total Videos
				Low Light	Dust	Rain	Fog	
Thermal	FLIR T650sc	Static Camera	Flat Cluttered Background	12	7	3	6	**28**
			Dynamic Background	8	8	5	5	**26**
		Motion Camera		3	1	0	2	**6**
Total Number of Videos				**23**	**16**	**8**	**13**	**60**
Total Time Duration				**44m46s**	**32m30s**	**16m19s**	**24m25s**	**1h58m**

the signal can be lost along the way, which produces blurry flat regions. The thermal infrared camera produces an image according to the differences in the omitted thermal radiation between an object and the background. If the background emits the same amount of thermal radiation as objects, e.g., a cluttered background, the foreground and background regions will be indistinguishable. We investigated the performance of a perceptual discrimination salient-feature-based methodology on a flat cluttered background, as shown in Fig. 2. The sample frames are collected from our TU-VDN dataset with a flat cluttered background. The pixel values of the background region in Fig. 2(a) and of the foreground object region in Fig. 2(b) are similar and vary smoothly; hence, the background and foreground true-positive pixel intensity values cannot be properly categorized, thereby resulting in incorrect interpretations. The main difficulty that is faced by well-known feature descriptors [35, 36] on such flat cluttered regions is homogenous neighbouring pixel intensity values. In Fig. 2(a), we have investigated a background-based local flat region where each neighbouring pixel similarity pattern (B_s) is computed using the centre pixel, where is marked as 'x'. In Fig. 2(b), we have also investigated a foreground-object-based local flat region that is cluttered with the background region. The foreground-region-based similarity pattern (F_s) has 6 matches out of 8 with the background-based similarity pattern (B_s), which could be categorized incorrectly as background. We have overcome over this challenge by increasing the robustness of existing local binary feature descriptors [35, 36], to obtain the ALWBP descriptor (details about this descriptor are presented in Section V). In Fig. 2(c), the ALWBP similarity pattern (A_s) is computed using a reference centre pixel, which is marked as '√' (details about the reference centre pixel are presented in Subsection A.2). As a result, the foreground similarity pattern (A_s) has 3 matches out of 8 with the background pattern (B_s), which is sufficiently discriminative to be correctly categorized as foreground.

IV. Related Work

Over decades, the object detection methods that have been used for visual frames, have also been used for thermal frames. The main objective of these methods is to determine whether a specified pixel intensity value is a true positive or not. In change detection, multiple strategic approaches have been applied: density-based [17, 18], sample-consensus-based [19, 20, 37], spatial-feature-extraction-based [35, 36, 41, 42, 43, 44], and fusion-based [21, 22] approaches. A prominent parametric method, namely, Gaussian mixture models (GMM), which was proposed by Stauffer *et al.* [17], typically performs adequately against shadowy multimodal background regions. Each background pixel is modeled using a mixture of Gaussian probability density functions via an iterative update rule. Another density-based estimation method, namely, kernel density estimation (KDE), which was introduced by Elgammal *et al.* [18], has been successfully applied in background segmentation. KDE is a non-parametric model that is used to estimate background probability density functions directly from local intensity observations. More flexible variations of GMM [38, 39] and KDE [40] have also been proposed over the years to improve the convergence rate.

The background-sample-based strategy was introduced by Wang *et al.* [37]. The sample consensus (SACON) is defined at each pixel according to the N most recent pixel intensity samples. However, most methods of this type are unable to model long-term periodic events because their observations are based on a first-in, first-out strategy. To overcome this problem, a random observation replacement strategy was introduced in [19, 20] into the background models. In [20], Droogenbroeck *et al.* proposed another non-parametric method, namely, visual background extractor (ViBE), which utilizes a random approach to update background pixels and diffuse their current pixel values into neighbouring pixels. The main drawback of the ViBE method is that it follows a global fixed parameter strategy for model maintenance, which faces the problem of dynamic variations of real-world scenes. In this case, Hofmann *et al.* [19] proposed a feedback scheme, namely, pixel-based adaptive segmenter (PBAS), for monitoring the background dynamics at the pixel level through adaptive state variables.

All these pixel-based models characterize pixels according to only colour intensity values. Colour intensities reflect the visual perception properties and often ignore part of the spatial information between adjacent pixels. Spatial-based feature extraction is beneficial in the typical cases where the foreground object texture is similar to the background's texture in terms of pixel intesities, especially in thermal video sequences. Heikkila *et al.* [35] were the first to explore the use

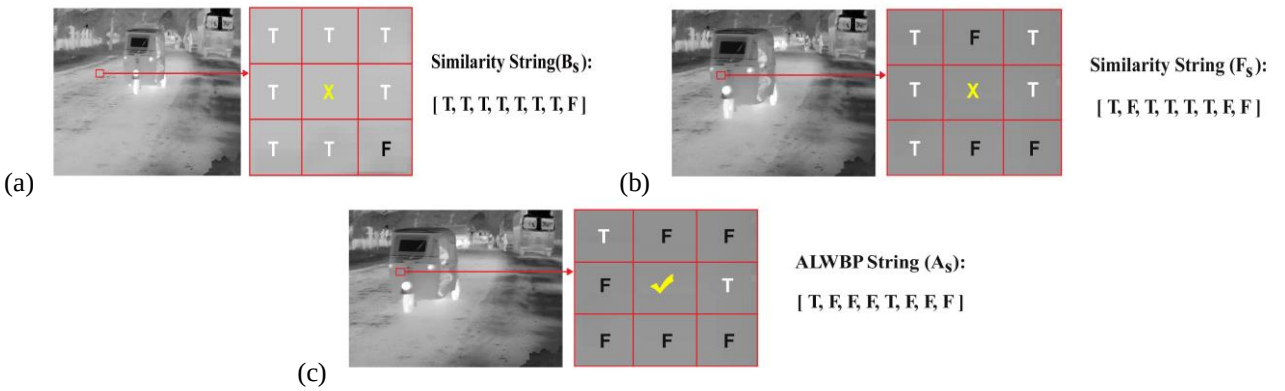


Fig. 2. Outline of the salient-feature-based methodology over a flat cluttered background. (a) Background flat region. Each neighbouring pixel similarity pattern (B_s) is computed using the center pixel (marked as 'x'); (b) Foreground object flat region. The foreground string (F_s) has 6/8 matches with the background similarity string (B_s), which could be categorized as background (incorrectly); (c) Foreground object flat region. The ALWBP descriptor (A_s) is computed using a randomly selected background sample (marked as '√') as a reference center pixel. The foreground string (A_s) has 3/8 matches with the background string (B_s), which is categorized as foreground (correctly).

of the local binary pattern (LBP) descriptor to improve the spatial awareness in background modeling. Their method used the LBP feature to handle illumination variations by comparing local pixel intensities. Since then, many modified variations of LBP have been proposed in the literature. In 2009, Heikkila *et al.* [41] introduced the centre symmetric LBP (CS-LBP) to further improve the computational efficiency. A new type of pattern features, namely, local binary similarity pattern (LBSP) features, were proposed by Bilodeau *et al.* [36], which are based on measure similarities instead of pixel comparisons and decrease the frequency of false classifications. Tan *et al.* [42] extended the binary pattern to a ternary pattern (LTP) by thresholding the pixel value differences to analyse the flat image regions. For detecting illumination changes at the pixel level, Liao *et al.* [43] presented a method, namely, scale-invariant LTP. A novel night time pedestrian detection method, namely, thermal-pixel-intensity-histogram-of-oriented-gradients (TPIHOG), was proposed by Baek *et al.* [44] for investigating the thermal and pixel intensities using the HOG descriptor.

In background modeling, numerous attempts have been made to combine the advantages of pixel-based and spatial-based approaches in the generation of the background model to control both the change detection sensitivity and the dynamic background scenes in practice. One very well-known method in this category is the self-balanced sensitivity segmenter (SuBSENSE), which was proposed by St-Charles *et al.* [21], where LSBP [36] features and the PBAS [19] feed-back model are combined to improve the spatiotemporal sensitivity. The authors also extend their work with a few general improvements in a new method: local binary similarity segmenter (LOBSTER) [22].

Several other well-known background subtraction methods for moving object detection are available in the literature. Maddalena *et al.* [45] presented a self-organizing artificial neural network for background subtraction (SOBS) for handling gradual illumination variations and camouflage. The Codebook methods of [46, 47] represent cluster observations as code words and store them in local dictionaries. The first eigenvalue-decomposition-based background model was proposed by Oliver *et al.* [48]. For more surveys on foreground segmentation, one can consult the many review papers [49, 50] that have been published.

V. Proposed Methodology

Most methods that are used for foreground object segmentation in video sequences that were captured by thermal or visual cameras are composed of three modules: maintenance of the background model, the current frame, and feature extraction. For thermal cameras, the incongruence between a background model and the current frame is not typically indicative for all types of objects. The only advantages of thermal cameras are that the captured images are not influenced by illumination and shadows and a pedestrian can be clearly distinguished as a foreground object due to its temperature absorbance. Other foreground objects, such as moving vehicles, that are comprised of several body components, such as wheels and headlights are visible, while the remaining components have similar texture to the as background. However, finding a satisfactory reference or background model for background subtraction is difficult when there are several real-time objects in thermal frames.

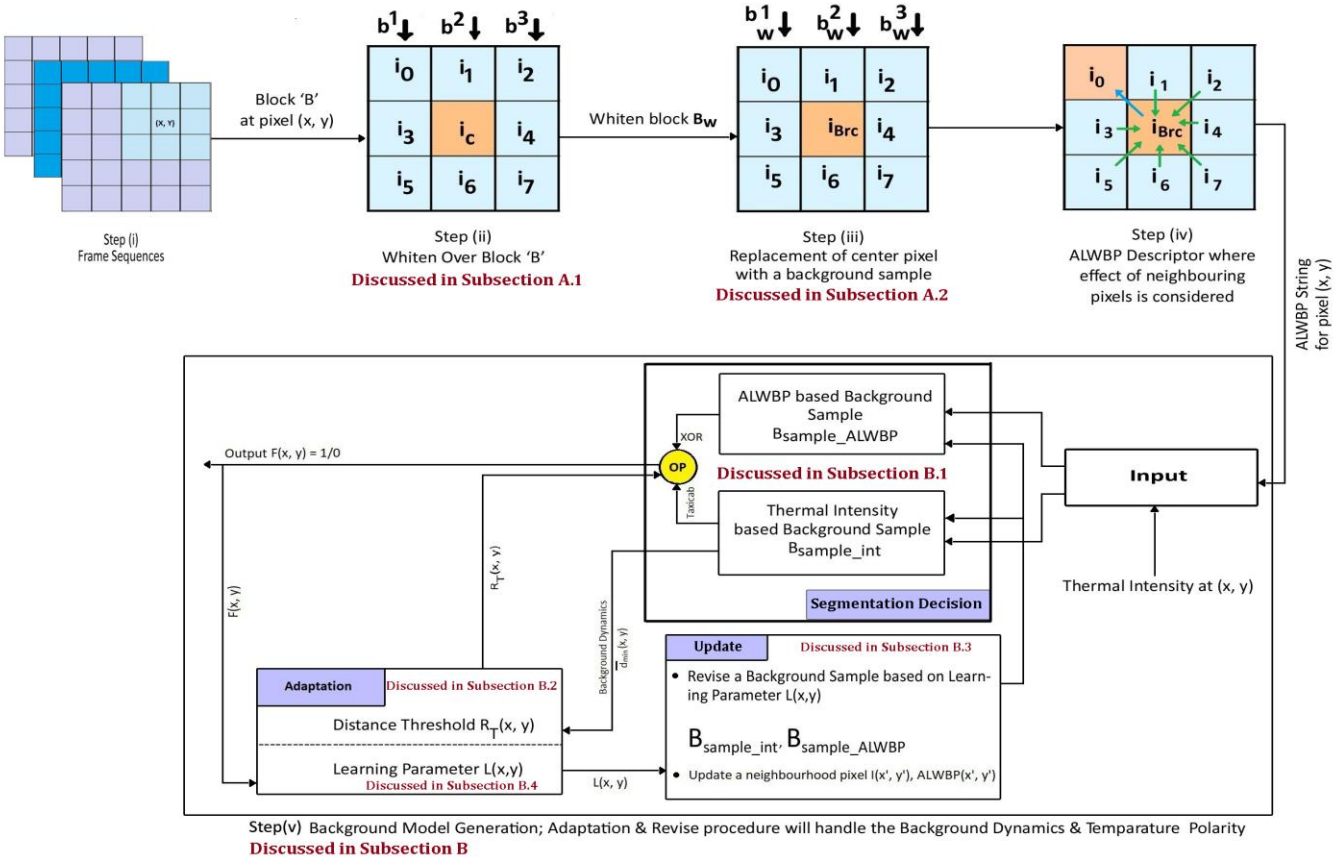


Fig. 3. Proposed system pipeline for background segmentation from thermal video sequences in adverse weather conditions.

In this paper, we present a satisfactory background segmentation model that uses the novel Akin-based Local Whitening Boolean Pattern (ALWBP) salient features; a pipeline of the system is shown in Fig. 3. It handles flat cluttered regions in thermal frame sequences and increased false-negative ratios. The model is inspired by pixel-level [19] and spatiotemporal-level [21] methods because LBP [35] or LBSP [36] features are not robust to flat cluttered regions when neighbouring pixels are similar. The overall system pipeline of the proposed background segmentation method is the combination of an ALWBP feature descriptor and a background model generation. The pipeline has been described briefly as follows:

Part 1: ALWBP Feature Descriptor
Step (i): The video clips are converted in frame sequences and extracted the local blocks over each pixel position.
Step (ii): We used well-known transformation *whitening* to mitigate the effect of similar correlation of homogeneous neighbouring pixel intensity values over the local block. A detail about this step is discussed in Subsection A.1.
Step (iii): The center pixel of local block is replaced by a randomly chosen background sample as reference centre pixel. It is substantial discriminative power even in homogeneous neighbouring pixels. A detail about this step is discussed in Subsection A.2.
Step (iv): The ALWBP descriptor is also described in Subsection A.2.
Part 2: Background Model Generation
Step (v): The part 2 collectively represents a generation of background model where both spatial-level and pixel-level features are represented as ALWBP Boolean patterns and thermal intensities respectively as inputs. It is consist of three sub steps: decision for segmentation, adaptation of parameters and updating the background samples.
Decision for Segmentation: To match each thermal pixel intensity or ALWBP Boolean features with background integer or ALWBP features, we have used the taxicab geometry distance or XOR logical operation via a *pixel wise dynamic threshold* or *Hamming distance threshold*. A detail about these operations is discussed in Subsection B.1.
Adaptation of Parameters: The adaptation of per-pixel *dynamic threshold* and *learning parameter* are also discussed in Subsection B.2 and B.4. For highly dynamic areas, the threshold value should be high to prevent incorrect classifications as foreground and be low for static areas.
Updating the Background Samples: In Subsection B.3, the updating of background pixels is explained. Thermal background intensity changes like in the first appearance of an object, waves of water layers, and shaking trees are considered in this section.

A. Video Saliency Feature

Salient-feature-based object detection has recently increased in popularity in computer vision research [1, 4]. According to the types of input, there are two categories of saliency models, namely, static and dynamic. The static models take still images as input and the dynamic models operate on video sequences. In this paper, we aim at detecting salient moving object regions in video scenes. The invention of salient features in outdoor adverse atmospheric videos is a highly challenging problem due to the complications that are encountered under the loss of contrast and motion information. Therefore, we propose a novel salient features for each pixel, namely, the *Akin based Local Whitening Boolean Pattern (ALWBP),* which is presented in Algorithm 1 and described as follows:

LBP [35] and LBSP [36] are well-performing and fast local feature descriptors, which are effective in analysing textures. Existing LBP and LSBP descriptors have the following disadvantages: (i) LBP only considers differences between the centre and each neighbouring pixels and (ii) LSBP considers the similarity between the centre and each neighbouring pixel, but not the effect of neighbouring pixels on the current similarity between the considered centre and neighbouring pixel. These methods are illumination invariant but not robust against low-frequency flat regions and smooth backgrounds or cluttered backgrounds, which has been discussed in the problem definition in Section III. These feature descriptors have difficulties on flat cluttered regions due to the homogeneous neighbouring pixel intensity values. Suppose we are extracting features on a thermal flat cluttered region block $B \in \Re^{n\times n}$ (B consists of vectors $b^i \in \Re^n$ for $1 \le i \le n$) where the values of adjacent pixels are highly correlated. B is a 3x3 block and each column is a set of three pixel values. Each 3x1 column vector is considered as feature vector b^i. Therefore, block B contains of three feature samples. It is necessary to pre-process each b^i such that the correlation values are lower between adjacent pixels. A very well-known approach is to *whiten* each b^i in the direction of pixel variations that are perpendicular to each other, such that they will have lesser correlation with unit variance [51].

A.1 Whitening Over a Local Block: To more formally identify the directions of $b^1, b^2, ..., b^n$, we compute the matrix covariance, namely, $\sum$, as follows:

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(b^i - \bar{b})(b^i - \bar{b})^T \quad (1)$$

The eigenvalue decomposition (EVD) can be used to analyse the covariance matrix $\sum$of $B \in \Re^{n\times n}$ as follows:

$$\Sigma = [u_1, u_2, ..., u_n][diag(\lambda_1, \lambda_2, ..., \lambda_n)][u_1, u_2, ..., u_n]^T \quad (2)$$

where u_1 is the principal vector, namely, the first eigenvector, of $\sum$; u_2 is the second eigenvector; and so on. These vectors are stacked to form an orthogonal matrix, which is denoted as U. Additionally, let $\lambda_1, \lambda_2, ..., \lambda_n$ be the corresponding eigenvalues; they form a diagonal matrix, which is denoted as D. To make our input vectors b^i less correlated with each other, we reflect the original data as follows:

$$b^i_{refl} = U^T b_i \quad (3)$$

Thus, $b^1_{refl}, b^2_{refl}, ..., b^n_{refl}$ will be less correlated and will satisfy one of our whitening properties. Since U is an orthogonal matrix, it satisfies the property $UU^T = U^T U = I$. Therefore, the reflected vector b^i_{refl} back to original data b^i can be computed via $Ub^i_{refl} = UU^T b^i = Ib^i = b^i$.

The unit variance properties of input vectors b^i are imposed by rescaling each reflected vector b^i_{refl} as follows:

$$b^i_{resl} = b^i_{refl} \Big/ \sqrt{\lambda_i + \varepsilon} \quad (4)$$

In the scaling step of Eq. (4), a small constant, namely, ε, is added to the eigenvalues to make the feature vectors numerically stable. Altogether, the whitening is defined as follows:

$$B_w = UB_{resl} \quad (5)$$

$$= \mathrm{U} \times \frac{b^i_{refl}}{diag(\sqrt{\lambda_i + \varepsilon})} \quad \text{using Eq.(4)}$$

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

$$= \mathrm{U} \times \mathrm{diag}((\lambda_i + \varepsilon)^{-\frac{1}{2}}) \times \mathrm{b}^{i}_{refl}$$

$$= \mathrm{UD}^{-\frac{1}{2}} B_{refl} \qquad \{\because \mathrm{B}_{refl} = [\mathrm{b}^1_{refl}, \mathrm{b}^2_{refl}, ..., \mathrm{b}^n_{refl}]\}$$

$$= \mathrm{UD}^{-\frac{1}{2}} U^T b^i \qquad \text{using Eq. (3)}$$

$$= \mathrm{UD}^{-\frac{1}{2}} U^T \mathrm{B} \qquad (6)$$

The matrix $\mathrm{B_w}$ of flat cluttered region block B is white, namely, its vectors $b^1_w, b^2_w, ..., b^n_w$ are less co rrelated and of unit variance as shown in *Proposition 1*. The covariance of matrix B_w satisfies the following identity property:

$$E\{B_w B_w^T\} = I \qquad (7)$$

Proposition 1: If the whitened feature vectors $B_w = \{b^1_w, b^2_w, ..., b^n_w\}$ follow a joint Gaussian distribution, they are independent.

Proof. Inspired by [52], we demonstrate that the feature vectors $b^1_w, b^2_w, ..., b^n_w$ of white matrix B_w about the pixels of a local flat region block follow a bivariate normal distribution. From the definition of a Gaussian orthogonal ensemble (GOE), the covariance of B_w is identity I, namely, the matrix is real and symmetric, and the probability density $GOMD[\sigma, n]$ represents a Gaussian orthogonal matrix distribution (GOMD) with matrix dimension $\{n \times n\}$ and scale parameter σ. In other words, the entries of matrix $I = I_{ij}$ are jointly proportional to a Gaussian with unit variance $\sigma^2 = 1$ and uncorrelated $\rho = 0$ [52].

If the joint distribution between any two feature vectors, such as b^1_w and b^2_w, can be written as the product of non-negative functions, feature vectors b^1_w and b^2_w will be independent as follows:

$$f(b^1_w, b^2_w)$$

$$= \frac{1}{2\pi\sigma_{b^1_w}\sigma_{b^2_w}\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(b^1_w-\mu_{b^1_w})^2}{\sigma^2_{b^1_w}} - \frac{2\rho(b^1_w-\mu_{b^1_w})(b^2_w-\mu_{b^2_w})}{\sigma_{b^1_w}\sigma_{b^2_w}} + \frac{(b^2_w-\mu_{b^2_w})^2}{\sigma^2_{b^2_w}}]\}$$

$$= \frac{1}{2\pi\sigma_{b^1_w}\sigma_{b^2_w}\sqrt{1-0^2}} \exp\{-\frac{1}{2(1-0^2)}[\frac{(b^1_w-\mu_{b^1_w})^2}{\sigma^2_{b^1_w}} - \frac{2\times 0\times(b^1_w-\mu_{b^1_w})(b^2_w-\mu_{b^2_w})}{\sigma_{b^1_w}\sigma_{b^2_w}} + \frac{(b^2_w-\mu_{b^2_w})^2}{\sigma^2_{b^2_w}}]\}; \text{ set } \rho = 0$$

$$= \frac{1}{2\pi\times 1\times 1} \exp\{-\frac{1}{2}[\frac{(b^1_w-\mu_{b^1_w})^2}{1^2} + \frac{(b^2_w-\mu_{b^2_w})^2}{1^2}]\}; \qquad \text{set } \sigma^2_{b^1_w} = 1, \sigma^2_{b^2_w} = 1$$

$$= \frac{1}{2\pi}\exp\{-\frac{1}{2}[(b^1_w-0)^2 + (b^2_w-0)^2]\}; \qquad \text{set } \mu^2_{b^1_w} = 0, \mu^2_{b^2_w} = 0$$

$$= \frac{1}{2\pi}\exp\{-\frac{1}{2}[(b^1_w)^2 + (b^2_w)^2]\}$$

$$= \exp\{-\frac{(b^1_w)^2}{2}\}\exp\{-\frac{(b^2_w)^2}{2}\}; \qquad \text{value of } \frac{1}{2\pi} \text{ is negligible}$$

$$= \mathrm{f}(b^1_w)\mathrm{f}(b^2_w) \rightarrow \text{two independent functions}$$

Therefore, the relation between Proposition I and Eq. (7) is that the feature vectors from transformed white matrix B_w are less correlated as well as independent because only making of less correlation does not mean independence.

A.2 ALWBP Descriptor: We have obtained a local flat region of pixels via Eq. (6) that are less correlated, and used this region to generate Akin-based local Boolean pattern (ALWBP). The term *Akin* indicates a most appropriate similar neighbouring pixel to a centre reference pixel that has more analogous characteristics than the other neighbouring pixels. Unlike the traditional LBP [35] and LBSP [36] approaches, which calculates the difference and similarity, respectively, between two pixel values (a centre pixel and a neighbouring pixel), the ALWBP approach considers the effect of other neighbouring pixel values. This Akin-based concept is termed *Akinity* [53] and is described in Fig. 4. *Brc* is a background intensity sample value at (x, y), which is the reference centre (*rc*) pixel. This differs from the approaches in [21, 22], where the reference centre pixel is imported from a previous frame intensity value. We have altered it because in flat regions, selecting the previous frame reference pixel as centre does not yield substantial discriminative power. The value of the reference centre from the background sample is selected randomly from N samples (regarding background samples, see Section V.B). While evaluating a '*candidate Akin neighbouring pixel*' for the '*centre pixel*', we consider other candidate Akin neighbouring pixels as competitors. Fig. 4 shows the *Akinity*, namely, $a(i_{x,y,Brc}, i_{x-1,y-1,p_0})$ from centre pixel $i_{x,y,Brc}$ to a candidate Akin neighbouring pixel $i_{x-1,y-1,p_0}$: Akin neighbouring pixel $i_{x-1,y-1,p_0}$ serves as the most similar candidate for centre pixel $i_{x,y,Brc}$, while other candidate Akin neighbouring pixels i' will compete for centre pixel $i_{x,y,Brc}$. Via this approach, we can analyse the similarity between two pixels more intensively than between other neighbouring pixels. The *Akinity* 'a' at location $(i_{x,y,Brc}, i_{x-1,y-1,p_0})$ can be calculated via the following formula:

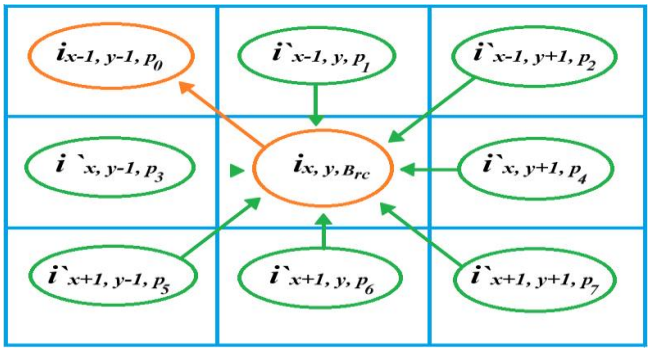


Fig. 4. Akinity $a(i_{x,y,Brc}, i_{x-1,y-1,p0})$ from center pixel $i_{x,y,Brc}$ to candidate Akin pixel $i_{x-1,y-1,p0}$.

$$a(i_{x,y,Brc}, i_{x-1,y-1,p_0}) = sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) - \min_{i' \neq i}\{sim(i_{x,y,Brc}, i')\}... \quad \text{if} \quad sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) < \mathrm{T_S}$$

$$= sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) + \min_{i' \neq i}\{sim(i_{x,y,Brc}, i')\}... \quad \text{if} \quad sim(i_{x,y,Brc}, i_{x-1,y-1,p_0}) \geq \mathrm{T_S} \qquad (8)$$

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

How much higher is the similarity score of a candidate Akin neighbouring pixel $i_{x-1,y-1,p_0}$ than those of the other competing candidate Akin neighbouring pixels i'? To answer this, we have subtracted the largest of the similarities among the competing candidate Akin neighbouring pixels i' with centre pixel $i_{x,y,Brc}$. At this point, we impose a condition: if the similarity between $i_{x,y,Brc}$ and $i_{x-1,y-1,p_0}$ is less than a similarity threshold, namely, T_s, (which is set to 0.2 in this paper), the value will be subtracted; otherwise it will be added. Hence, if there is a more correlated value even after the whitening process, the similarity will be increased, and if there is a slightly uncorrelated value between centre and an Akin neighbouring pixel, the similarity will be decreased. In a same manner, the *Akinity* will be estimated for remaining neighbouring pixels, namely, $p_1, p_2, ..., p_7$.

Since the *Akinity* is estimated among a group of neighboring pixels with a centre point, in some circumstances, replicate values will be obtained, which is called oscillation of numerical values. It is important for them to be damped to avoid numerical oscillation. Each updated damped *Akinity* value is set to λ times its previous value plus $(1-\lambda)$ times its current *Akinity* value, as follows:

$$a_p(i_{Brc}, i_p) = \lambda \times a(i_{Brc}, i_{p-1}) + (1-\lambda) \times a(i_{Brc}, i_p)\ ;\ 0<p\le 7 \quad (9)$$

In our case, the damping factor (λ) value is 0.5. Furthermore, we define a lower bound and an upper bound, namely, $a_p^{lower} < a_p < a_p^{upper}$, such that the *Akinity* values are 0 and 1. Via this approach, we have tried to evaluate the discriminative nature, even in flat regions, which helped increase the number of true-positive values and decrease the number of false-negative values.

Now, the ALWBP descriptor Boolean string rule is presented as

$$ALWBP(x,y) = T \quad \text{if } a_p < relative\_tau\ ;\ 0 \le p \le 7$$
$$= F \quad \text{Otherwise} \quad (10)$$

where a_p corresponds to the estimated *Akinity* value of the p^{th} neighbour of the pixel at (x, y) in the current frame and $relative\_tau = E \times i_c$ is the new energy based threshold value estimate for the current centre pixel at (x, y). To capture the micro-texture in a smooth region, the spatial two-dimensional dependence matrix, which is known as the *grey-level co-occurrence matrix* (G), of thermal grey palette values is used with displacement vector $d = (dx, dy)$, where *dx=1* and *dy =1* [64]. The feature that measures the randomness of grey-level distribution is the energy, namely, E, which is defined using the grey-level co-occurrence matrix as follows:

$$Energy(E) = \sum_x \sum_y G^2[x,y] \quad (11)$$

The ALWBP salient feature descriptor is calculated in Algorithm 1. An example of an ALWBP simplified theory-based description is presented in Fig. 2 and a simulation result is presented in Fig. 5.

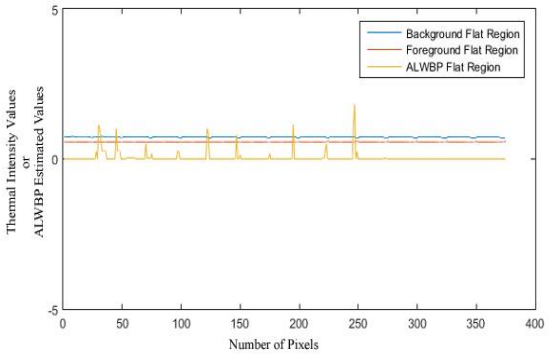


Fig. 5. The discriminative nature between 'raw pixel values of background and foreground region' and 'estimated pixel values of ALWBP descriptor' (Estimation process of ALWBP values are similar to LBP).

Algorithm 1 (ALWBP): Akin-based Local Whitening Boolean Pattern Generation

Input: A frame *F*, the energy (*E*) over frame *F*.
Output: A Boolean pattern string for each pixel in frame *F*.

1. **for** *x*: length(*F*,*1*)
1.1 **for** *y*: length(*F*,*2*)
1.1.1 B_{rc} = randomly select a background sample from *N* recent samples
1.1.2 *relative_tau* = *E*F(x,y)*
1.1.3 *B* = extract a 3x3 block at coordinate*(x,y)*
1.1.4 Initialize a matrix of size *B*, *a*←0
1.1.5 Whiten the block *B* *// to reduce the correlation between adjacent pixels*
1.1.6 $B_w(i_c,i_c) = B_{rc}$ *// center is replaced by a background sample that is used as the reference center pixel*
1.1.7 Estimate *Akinity* '*a*' over this whitened block B_w as
if $sim(i_{Brc},i_p)<T_s$ **then**
1.1.7.1 $a(i_{Brc},i_p) = sim(i_{Brc},i_p) - min\{ i_{Brc},i'_p \}$
1.1.8 **else**
1.1.8.1 $a(i_{Brc},i_p) = sim(i_{Brc},i_p) + min\{ i_{Brc},i'_p \}$
1.1.9 **endif**
1.1.10 Dampen the *Akinity* values to avoid numerical oscillations via
$a_p(i_{Brc},i_p) = \lambda * a(i_{Brc},i_{p-1}) + (1-\lambda) * a(i_{Brc},i_p)$
1.1.11 **if** $a_p<relative\_tau$ **then**
1.1.11.1 *ALWBP(x,y) = T*
1.1.12 **else**
1.1.12.1 *ALWBP(x,y) = F*
1.1.13 **endif**
1.1.14 **endfor**
1.1.15 **endfor**

B. Generating the Background Model via ALWBP (BM∪ALWBP)

To generate our non-parametric background model, we represent each background pixel using both spatial-level and pixel-level features, namely, ALWBP Boolean patterns and thermal intensities. To try to match each pixel from the current frame with background integer samples, we first compare the thermal pixel intensity values using the taxicab geometry to a pixel wise dynamic threshold R_T. Second, we compare the ALWBP Boolean patterns over 3x3 blocks on the current frame with background ALWBP samples via a Hamming distance threshold, which is denoted as H_T. Regardless of whether a pixel belongs to the background or foreground, both thermal intensities and ALWBP Boolean patterns are considered in our method, as in [21, 22]. The methods in [21, 22] are typically not robust against flat cluttered backgrounds, whereas our method focuses on this issue, along with other background subtraction problems such as thermal intensity

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

changes upon the first appearance of objects and dynamic backgrounds.

B.1 Pixel decision via samples of thermal intensity and ALWBP: Inspired by [19, 20], we develop a random sample consensus framework for modelling both long-term and short-term periodic events [21]. Each pixel intensity, namely, $I(x, y)$, is modeled by an array of N recently observed background intensity samples, namely, B_{sample_int} and ALWBP string samples, namely, B_{sample_ALWBP}.

$$B_{sample\_\text{int}}(x, y) = \{BInt_1(x, y), BInt_2(x, y), ..., BInt_k(x, y), ..., BInt_N(x, y)\} \in \Re^{N\times 1} \quad (12)$$

$$B_{sample\_ALWBP}(x, y) = \{BALWBP_1(x, y), BALWBP_2(x, y), ..., BALWBP_k(x, y), ..., BALWBP_N(x, y)\} \in S^{N\times 1} \quad (13)$$

For thermal scenes, N must be a small as possible to balance memory consumption and computational complexity (we set $\#N = 10$ in our case). Each of these samples is matched against its observation $I(x, y)$ or $ALWBP(x, y)$ at coordinate (x, y) on the current frame for classifying a pixel as foreground ($F(x, y) = 1$) or background ($F(x, y) = 0$) as follows:

$$F(x, y) = 1 \quad \text{if } \{\text{texicab}(\text{I}(x,y),\text{B}_{sample\_\text{int}}(x,y)) < \text{R}_T(x,y) \ \& \ \text{XOR}(\text{ALWBP}(x,y),\text{B}_{sample\_ALWBP}(x,y)) \le H_T\} < \text{Threshold}_{\min}$$
$$= 0 \quad \text{otherwise} \quad (14)$$

In Eq. (14), $F(x, y) = 1$ corresponds to a per-pixel output segmentation map; $R_T(x, y)$ is the per-pixel distance threshold at pixel (x, y), which should be high for highly dynamic areas and low for static areas; H_T is a fixed Hamming distance threshold (we set $\#H_T = 3$); and for classification, $Threshold_{\min}$ is the minimum number of matches with background samples in both the thermal intensity and ALWBP pattern, which is a fixed global parameter (we set $\#Threshold_{\min} = 2$) that balances the computational complexity and noise resistance [19,20, 37].

B.2 Per-pixel adaptation of the distance threshold (R_T): A dynamic distance threshold, namely, R_T is defined per-pixel at coordinates (x, y). For highly dynamic areas, $R_T(x, y)$ should be high to prevent incorrect classifications as foreground and it should be low for static areas. In a video sequence, there can be regions with waving of a water layer or trees in the wind, which will provide higher background dynamics and result in incorrect classifications of foreground objects. In addition, there can be regions with small to no changes, which provide low dynamic value. Therefore, the background dynamics, namely, $\bar{d}_{\min}(x, y)$, must be estimated, as inspired by [19].

In addition to saving arrays of the N recently observed background thermal intensity samples and ALWBP samples in the background maintenance, as in Eq. (12) and (13), we create another array, namely, $D(x, y)$ of minimum-distance samples between the current thermal pixel intensity and the background intensity samples as follows:

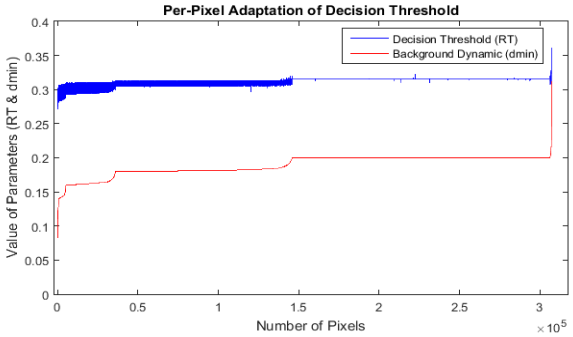


Fig. 6. Higher background dynamics (d_{min}) require faster threshold increments in the decision threshold (R_T) and the thresholds gradually decrease for in low background dynamic values.

$$D(x, y) = \{D_1(x, y), D_2(x, y), ..., D_k(x, y), ..., D_N(x, y)\} \quad (15)$$

$$D_k(x, y) = \min\{texicab(I(x, y), B_{sample\_\text{int}}(x, y))\} \quad (16)$$

To measure the background dynamics at pixel coordinate (x, y), the average of these minimum distance samples is calculated as follows:

$$\bar{d}_{\min}(x,y) = \frac{1}{N}\sum_{k=1}^{N} D_k(x,y) \quad (17)$$

The dynamic adaptation of distance threshold $R_T(x, y)$ via this measurement of the background dynamics is expressed as follows:

$$R_T(x, y) = (1 - R_{lr}) \times R_T(x, y) + R_{lr} \times \bar{d}_{\min}(x, y) \quad (18)$$

where R_{lr} is a fixed regulated controller rate for the distance threshold ($R_{lr} = 0.02$ in our case). In completely static regions or less dynamic background regions, namely, $\bar{d}_{\min} \cong 0$, the value of $R_T(x, y)$ will slowly decrease. In contrast, under increasing background dynamics, the distance threshold, namely, $R_T(x, y)$, approaches the product value of $R_{lr} \times \bar{d}_{\min}(x, y)$, which provides a robust, increasing threshold value. However, in dynamic regions, $R_T(x, y)$ initially slightly decreases by a factor of $(1 - R_{lr})$ and subsequently rapidly increases by a factor of R_{lr} as the value of $\bar{d}_{\min}(x, y)$ increases. In Fig. 6, the decision threshold R_T is plotted.

B.3 Updating the Background Model: To account for changes in the background, such as thermal intensity changes upon the first appearance of an object in the frame (as shown in Fig. 7), a waving water layer, and shaking trees, updating the background pixels in the background model, namely, B_{sample_int}, B_{sample_ALWBP}, is essential.

(a) (b)

Fig. 7. Thermal intensity changes upon the first appearance of an object in (a) a thermal frame and (b) the next frame in which the object enters for the first time.

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

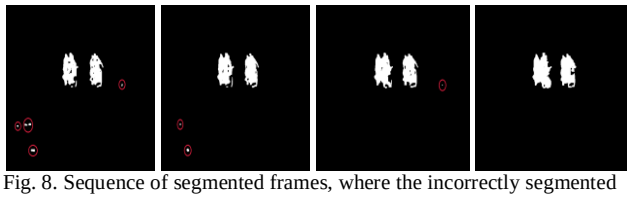


Fig. 8. Sequence of segmented frames, where the incorrectly segmented (marked by red circles) foreground pixels are gradually vanish in subsequent frames.

We have updated our background model via a similar approach to that in [19]. A pixel at coordinate (x, y) is updated to one of the background samples if and only if the pixel is categorized as background, namely, $F(x, y) = 0$. Hence, foreground pixels will be excluded from this update process. For a randomly selected index $k \in 1, 2, ..., N$, the corresponding background sample values, namely, $BInt_k(x, y)$ and $ALWBP_k(x, y)$, are replaced by the current intensity value, namely, $I(x, y)$, and $ALWBP$ pattern, namely, $ALWBP(x, y)$, respectively. At the same time, we also update a random sample that is selected from 8-neighbouring pixels: $I(x', y') \in N(I(x, y))$. The background model at this neighbouring pixel is replaced by its current intensity value, namely, $I(x', y')$, andpattern, namely, $ALWBP(x', y')$. Via this neighbouring pixel update process, wrongly classified foreground pixels are gradually incorporated into the background model, as shown in Fig. 8.

B.4 Per-pixel adaptation of the learning parameter (L): Every pixel, whether foreground or background, that is incorporated into a background sample also depends upon the learning parameter, namely, $L(x, y)$. A higher $L(x, y)$ value indicates that the pixel at (x, y) is more likely to be incorporated into the background model. Here, we omit the probability concepts for simplicity [19]. According to the adaptation of the learning parameter $L(x, y)$, pixels those pixels that are wrongly classified as foreground will be merged into background pixels. This strategy is formulated in Eq. (19) as follows:

$$L(x, y) = L(x, y) \times \{(1 - L_{lr} / \bar{d}_{\min}(x, y)) \times F(x, y) + ... \\ (1 + L_{lr} / \bar{d}_{\min}(x, y)) \times (1 - F(x, y))\} \quad (19)$$

where L_{lr} is a learning rate (L_{lr} = 0.02 in our case). The learning parameter of a pixel is decreases fast if the pixel belongs to the foreground, namely, if $F(x, y) = 1$, or a plus low dynamic background and slowly decreased in the case of a highly dynamic background. As a result, an incorrectly classified pixel will slowly be identified as background pixel. If a pixel belongs to the background, namely, if $F(x, y) = 0$, the second term in Eq. (19) (after '+') will increase the learning parameter value by $L_{lr} / \bar{d}_{\min}(x,y)$. Hence, a pixel is assumed by default to belong to the background and the learning rate will increase based on the value of $\bar{d}_{\min}$. A larger value of $\bar{d}_{\min}$ will slowly increase the learning parameter value, namely, $L(x, y)$, and small value of $\bar{d}_{\min}$ will rapidly increase it. Y. Wu *et. al.*[62] also suggested an similar approach which uses a progressive model and depends on the pre-set enlarging factor. The smaller enlarging factor indicates lower enlarging speed and more training time, but tends to result in a promising performance in the end. The larger enlarging factor which argues pseudo labeled candidate set will increase rapidly, but as a result it may not be reliable enough to train a robust CNN model. In our approach, the Eq. (19) is computed dynamically in nature. The variable $\bar{d}_{\min}(x,y)$ is dependent on the background dynamics for each pixel location.

Algorithm 2 (BM ⋃ ALWBP): Background Model using Akin based Local Whitening Boolean Pattern

Input: Total number N of frame sequences F^i for the generation of the corresponding background model

Output: Corresponding segmented frame sequences F^i

```
1.      for i : N number of frames
2.          Initialization:
                match←0
                k←0
                R_T←initialize randomly
                L←initialize randomly
                Threshold_min←2
3.          for x : length(F^i,1)
3.1             for y : length(F^i,2)
3.1.1               while(k<=N) do
3.1.1.1                 if [taxicab{I(x,y), BInt_k(x,y)}<R_T(x,y) &&…
                        {ALWBP(x,y) ⊕ BALWBP_k(x,y)}<H_T(x,y)]
3.1.1.1.1                   match = match + 1
3.1.1.2                 endif
3.1.1.3                 k = k + 1
3.1.2               endwhile
3.1.3               if (match<Threshold_min) then
3.1.3.1                 F^i(x,y) = 1
3.1.4               else
3.1.4.1                 F^i(x,y) = 0
3.1.5               endif
3.1.6               Adaptation of distance threshold R_T(x,y) based on
                    dynamic parameter d̄_min(x,y)
3.1.7               Adaptation of Learning parameter L(x,y) based on
                    dynamic parameter d̄_min(x,y) and F^i(x,y)
3.2             endfor
4.          endfor
5.      endfor
```

Background or foreground segmentation using both the ALWBP feature and the thermal intensity is presented in Algorithm 2. In summary, our approach performs effectively on thermal frame sequences. We (i) evaluated the discriminative performance of the ALWBP descriptor in most flat regions and (ii) developed a robust learning background model that is based on ALWBP and thermal intensity features and accounts for variations of the thermal intensities and background dynamics based on the adaptation of two dynamic thresholding parameters: $R_T(x, y)$ and $L(x, y)$.

VI. Experimental Evaluations

This section is divided into three parts. **First**, we evaluate the performance of our proposed method, namely, BM ⋃ ALWBP under our night dataset, namely, *TU-VDN* , which consists of four atmospheric conditions, namely, low-light, dust, rain, and fog, along with the key challenges of static and moving cameras and cluttered and dynamic backgrounds, which are scenarios that are typically encountered in practice. **Second**, a total of 14 background-model-based moving object detection methods are compared with our proposed method on our *TU-VDN* dataset. These state-of-the-art object detection methods are Vibe [20], Subsense [21], LOBSTER [22], PAWCS [54], FST [55], PBAS [19], Multicue [56], ISBM [57], MTD [58], VuMeter [59], KDE [18], MoG_V2 [17], Eigenbackground [48], Codebook [47], MSCL-FL [60] and MBS [61]. Most of these methods have been implemented in BGSlibrary. **Third**, we also assess our proposed method using a widely popular

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

change detection dataset: CDnet 2014 [9]. The CDnet 2014 dataset contains total fifty three videos from eleven video categories with four to six videos sequences in each category, namely, bad weather, low frame rate, night videos, ptz camera, thermal, shadow, intermittent object motion, camera jitter, dynamic background, and baseline turbulence. The bad weather category consists of only four videos under several snow situations. In contrast, our TU-VDN dataset contains of total sixty night videos to explore several adverse weather conditions, namely, foggy, dusty, rainy, and low-light. With CDnet 2014 dataset, the evaluation is conducted on three selected categories, namely, *Thermal*, *badWeather*, and *Night*, because these categories are related to our dataset. The results are reported with respect to ten performance metrics, namely, recall, specificity, miss rate, fall out, precision, error rate, accuracy, jaccard index (JI), F_β-score, and MCC. To better assess the overall performance and compare the performances among state-of-the-art methods, we considered the following metrics: accuracy, F_β-score and MCC.

A. *Evaluation on the TU-VDN dataset*

To demonstrate our key contributions via the analysis of our dataset using the proposed technique, we present the performance evaluation in Tables II through eight in terms of segmentation evaluation metrics: *fallout* or *FPR*, *miss rate* or *FNR*, *specificity*, *accuracy*, *precision*, *recall*, *JI*, *F_1-score*, and *MCC*. In Table III, we also present a comparative performance evaluation on the *TU-VDN* thermal night dataset.

The complete results of our proposed method on the created dataset are listed in Table II. We evaluated the performance in overcoming the following key challenges: (i) **flat cluttered background:** The overall performance against a cluttered background under in *foggy conditions* is promising. The MCC scores are correlated with the F_1-scores, and JI is not a reliable metric for our dataset. The same correlation is also observed between the specificity and accuracy metric values. The F_1-score and MCC metric values are affected by the low precision value, whereas our proposed method's recall value is very high. Under *rainy conditions*, we realize the highest recall value because as the aerosol size increases (the radii of the rain droplets are on the order of microns), less scattering is observed. Therefore, there is less loss of contrast, which reduces the false-negative classification rate for salient objects areas [3]. The miss rates are satisfactory metrics; (ii) **dynamic background:** The dynamic background scenarios are well handled by our proposed method, as exemplified by the fall-out values, namely, the false-positive rates are very low. Therefore, the precision values are high and produce a balanced factor with recall values, which results in better F_1-score and MCC compared to a cluttered background. Our method well handles the foggy conditions, even under higher aerosol density; and (iii) **camera motion:** According to the segmentation results, camera motion poses the biggest challenge due to camera vibrations that occur in parallel with vibrations of the vehicle on which the camera is mounted. The state-of-the-art methods also fail to properly segment salient

TABLE II: RESULTS OF OUR METHOD ON THE TU-VDN DATASET. COLOURS ARE USED TO INDICATE KEY CHALLENGE METRIC VALUES – *ORANGE* FOR A CLUTTERED BACKGROUND, *PURPLE* FOR A DYNAMIC BACKGROUND, AND *BLACK* FOR CAMERA MOTION.

Atmospheric Conditions	Key Challenges		Fall Out	Miss Rate	Specificity	Acc.	Precision	Recall	JI	F_1-Score	MCC
Low Light	Static Camera	Flat Cluttered Background	0.0104	0.1838	0.9896	0.9844	0.6271	0.8162	0.5545	0.6963	0.7008
		Dynamic Background	0.0051	0.2309	0.9949	0.9937	0.6387	0.7691	0.6147	0.6979	0.7018
	Camera Motion		0.1022	0.3581	0.8978	0.8746	0.4518	0.6419	0.3464	0.5067	0.4662
Dust	Static Camera	Flat Cluttered Background	0.0279	0.2296	0.9721	0.9600	0.6060	0.7704	0.4998	0.6530	0.6501
		Dynamic Background	0.0037	0.2648	0.9963	0.9926	0.7710	0.7352	0.6027	0.7473	0.7464
	Camera Motion		0.0885	0.5991	0.9115	0.8571	0.4653	0.52009	0.3374	0.4918	0.4306
Rain	Static Camera	Flat Cluttered Background	0.0136	0.1485	0.9864	0.9843	0.4918	0.8515	0.4501	0.6122	0.6341
		Dynamic Background	0.0039	0.1545	0.9912	0.9911	0.5988	0.8433	0.5523	0.7003	0.7123
Fog	Static Camera	Flat Cluttered Background	0.0036	0.2128	0.9964	0.9937	0.7007	0.7872	0.5741	0.7215	0.7294
		Dynamic Background	0.0013	0.1391	0.9987	0.9982	0.7268	0.8609	0.6417	0.7686	0.7797
	Camera Motion		0.2384	0.5891	0.7616	0.7732	0.4298	0.5119	0.3087	0.4896	0.4541

TABLE III: COMPARISON IN TERMS OF F_1-SCORE, MCC, AND ACCURACY PERFORMANCE MEASURES OF 15 METHODS ON THE TU-VDN DATASET. COLOURS ARE USED TO INDICATE THE RANKINGS OF THE METHODS – GREEN FOR THE BEST-PERFORMING METHODS, BLUE FOR THE SECOND BEST-PERFORMING METHODS, AND RED FOR THE WORST METHODS.

State-of-the-art Methods, Year	Low Light			Dust			Rain			Fog		
	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.
Proposed Method	0.6555	0.6658	0.9891	0.7002	0.6983	0.9763	0.6962	0.6832	0.9877	0.7451	0.7456	0.9960
Vibe, 2011	0.5738	0.5954	0.9907	0.5565	0.5823	0.9740	0.7307	0.7379	0.9877	0.6844	0.7119	0.9921
Subsense, 2015	0.4812	0.5091	0.9837	0.5405	0.5679	0.9722	0.6112	0.6171	0.9788	0.8204	0.8218	0.9969
LOBSTER, 2014	0.5244	0.5413	0.9890	0.4967	0.5346	0.9688	0.5658	0.5949	0.9793	0.6649	0.6769	0.9943
PAWCS, 2016	0.3155	0.3505	0.9870	0.2322	0.2872	0.9656	0.6628	0.6752	0.9875	0.5176	0.5559	0.9945
FST, 2014	0.4061	0.4509	0.9523	0.2360	0.2564	0.7189	0.6760	0.6938	0.9892	0.5173	0.5659	0.9677
PBAS, 2012	0.5668	0.5803	0.9901	0.4033	0.4311	0.9547	0.7035	0.6893	0.9823	0.6936	0.7086	0.9952
Multicue, 2012	0.4961	0.5314	0.9798	0.6166	0.6345	0.9714	0.5511	0.5912	0.9717	0.5511	0.5913	0.9717
ISBM, 2011	0.3594	0.4064	0.9654	0.2191	0.2443	0.7040	0.2773	0.3387	0.9312	0.4951	0.5456	0.9501
MTD, 2010	0.4656	0.4927	0.9814	0.4709	0.4772	0.9729	0.4843	0.4948	0.9802	0.5561	0.5640	0.9928
VuMeter, 2006	0.3336	0.3882	0.9862	0.2171	0.2578	0.9667	0.6225	0.6373	0.9814	0.3071	0.3698	0.6083
KDE, 2000	0.3845	0.3996	0.9691	0.2689	0.2978	0.9453	0.6198	0.6315	0.9606	0.5332	0.5522	0.9938
MoG_V2, 1999	0.3119	0.3486	0.9854	0.2208	0.2799	0.9673	0.3532	0.3963	0.9651	0.2117	0.2509	0.9928
Eigenbackground, 2000	0.3695	0.4190	0.9673	0.3603	0.3996	0.9184	0.2120	0.1761	0.6499	0.3413	0.3734	0.9629
Codebook, 2004	0.3093	0.3623	0.8848	0.2066	0.2245	0.6807	0.2736	0.3958	0.9021	0.2319	0.3299	0.9111

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/TCSVT.2019.2926164, IEEE Transactions on Circuits and Systems for Video Technology

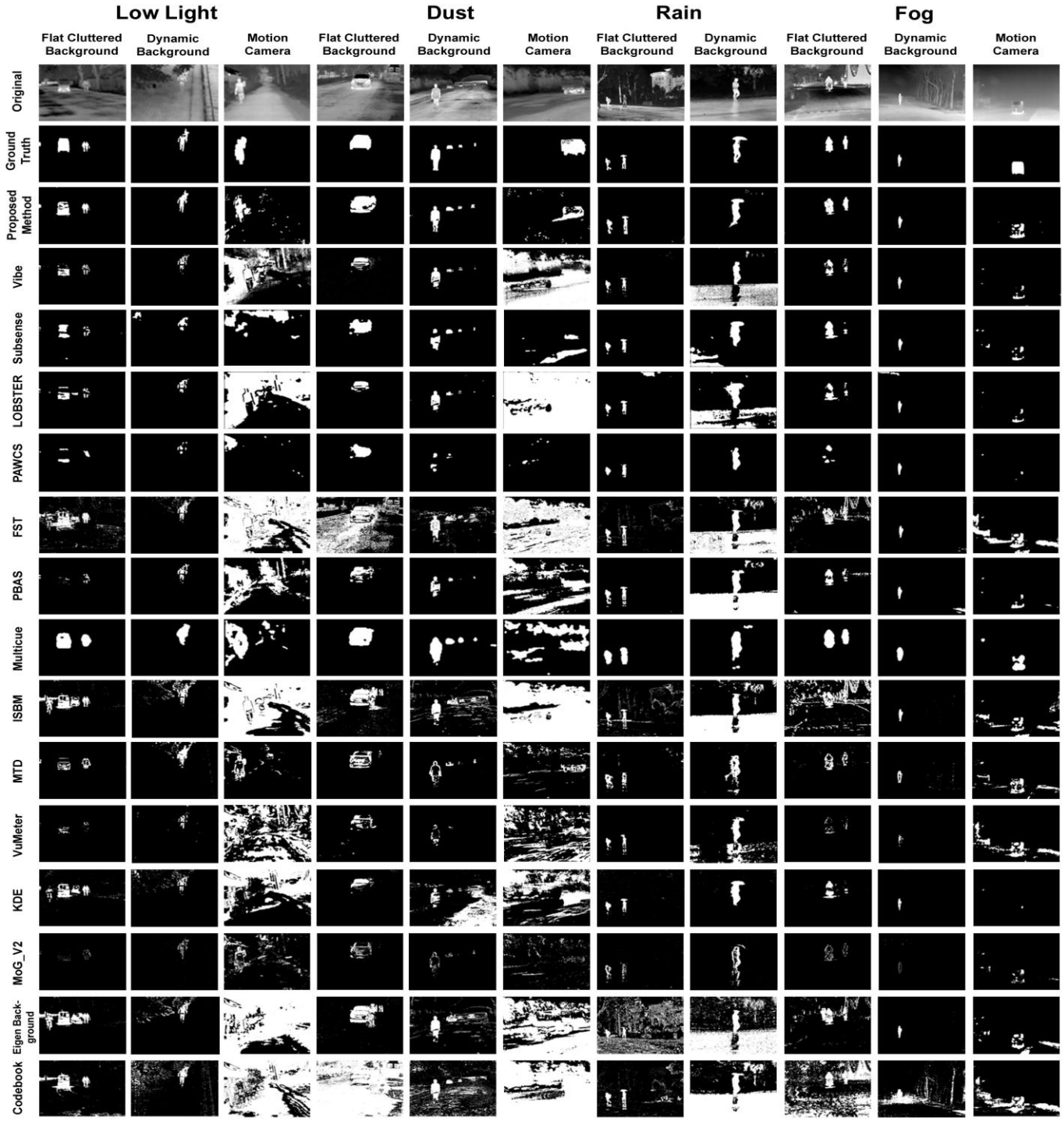
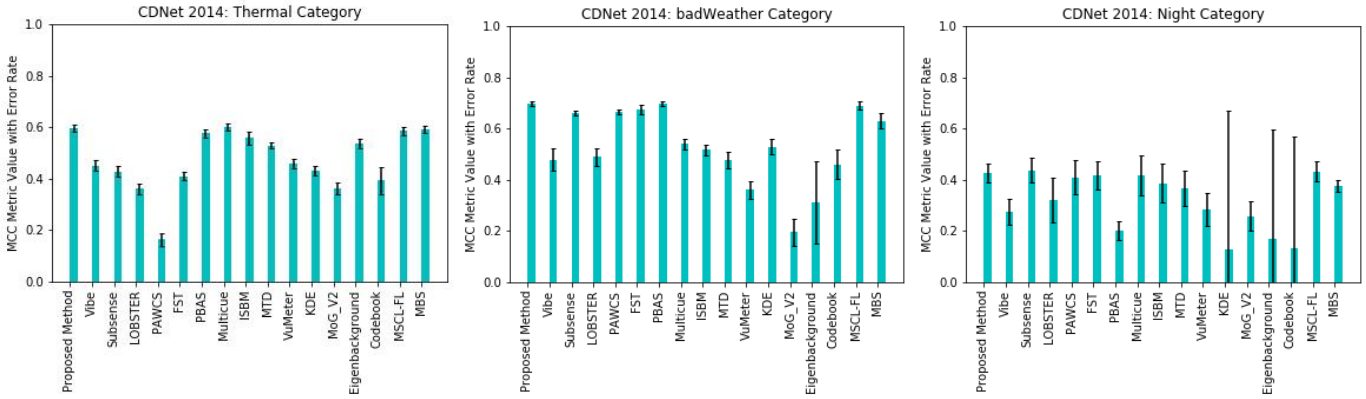


Fig. 9. Typical segmentation results for several key challenges under various atmospheric conditions in our created night time dataset. Row (1) shows input frames, row (2) shows the ground truth, row (3) shows the BMUALWBP results, row (4) shows the ViBe results, row (5) shows the Subsense results, row (6) shows the LOBSTER results, row (7) shows the PAWCS results, row (8) shows the FST results, row (9) shows the PBAS results, row (10) shows the Multicue results, row (11) shows the ISBM results, row (12) shows the MTD results, row (13) shows the VuMeter results, row (14) shows the KDE results, row (15) shows the MoG_V2 results, row (16) shows the Eigenbackground results, and row (17) shows the Codebook results.

object areas, as shown in Fig. 9. The fall-out and miss rate values are also very high. Thus, the camera-motion-based metric values are excluded from the estimation in the average performance comparisons in Table III.

According to Table III, the proposed model is robust in comparison to relevant classical and recently proposed state-of-the-art alternative methods. Under *low-light conditions*, our proposed method realizes an approximately 8% relative F_1-score and MCC improvement over the second best-performing method, namely, *Vibe*; *Codebook* and *MoG_V2* yield the poorest results. In terms of accuracy, *Vibe* and *Multicue* outperform the proposed method. Under *dusty conditions*, our method exhibits satisfactory performance according to all metrics, with nearly 6% and 8% relative F_1-score and MCC improvements over the second best-performing method, namely, *Multicue*, and 0.23% increased accuracy relative to *Vibe*. As usual, *Codebook* yields the poorest results. Under *rainy conditions*, *Vibe* and *PBAS* yield the most promising metric values. Our method exhibits the third best-performance and the *Eigenbackground* method performs the worst. Under *foggy conditions*, *Subsense* yields the best metric values, followed by our method, whereas the *MoG_V2* method yields poor results. To provide a better visual understanding of the categorization results, typical segmentation results are shown



(a) (b) (c)

Fig. 10. Comparative analysis on the CDnet-2014 change detection dataset over three categories: (a) the Thermal sequence, (b) the badWeather sequence, and (c) the Night sequence.

in Fig. 9 under various atmospheric conditions, along with key challenges.

B. *Evaluation on the CDnet 2014 dataset*

Our TU-VDN dataset is related to a few categories from the CDnet 2014 dataset. These scenarios are highly complicated due to low camera resolution. We analysed these categories separately via our proposed method and compared the results with those of all existing well-known state-of-the-art background subtraction methods. In Fig. 10 (a, b, c), we present bar graphs of the MCC metric values with error rates that were obtained by processing the *Thermal*, *badWeather*, and *Night* categories in the CDnet2014 dataset using seventeen state-of-the-art approaches, including our proposed approach. In the *thermal* category, our method realizes the second-best MCC value of 0.5943 with a marginal MCC difference from *Multicue* of 0.6001 and the lowest error rate of 0.0142. Whereas MBS and MSCL-FL performances set as third- and fourth-best methods with MCC values of 0.5921 and 0.5843. The *PBAS* and *Eigenbackground* methods also yielded satisfactory results, whereas *PAWCS* yielded the poorest results. In the case of *badWeather*, the proposed method realizes the best MCC value of 0.6971 with a slide variation from *PBAS* of 0.6961 and *PAWCS* realizes the lowest error rate of 0.0095. *MSCL-FL* and *PAWCS* yield the third- and fourth-best results and *MoG_V2* the poorest. Last, on the most complicated night visual camera sequence, *Subsense* and *MSCL-FL* yields a promising MCC values, and MBS and our proposed method realizes the lowest error rates respectively. The remaining methods exhibit very poor performances. Among these three categories, the performance on the *badWeather* scenarios is superior to those on the *Thermal* and *Night* sequences.

VII. Conclusion

We have described briefly our newly created night video dataset, namely, TU-VDN, for moving object detection in thermal infrared images. The dataset consists of degraded atmospheric night outdoor scenes under low-light, dusty, rainy, and foggy conditions. We also presented a video salient-feature-based background segmentation technique that uses both spatial features and thermal intensity for the robust investigation of thermal frames. We summarize the findings regarding this proposed technique as follows: (a) it handles various key challenges in thermal outdoor scenes, such as dynamic background, flat cluttered background, and thermal intensity adjustment during the maiden appearance of a moving object in the video sequence; (b) in terms of accuracy, F_1-score, and MCC, the results of the comparative experiments on the TU-VDN dataset has demonstrated the superior performance of our proposed method; (c) the results of our analysis on the CDnet-2014 dataset over the night, thermal, and badWeather category sequences have also demonstrated the superior performance of the approach in terms of MCC value and error rate.

References

[1] R.T. Tan, "Visibility in bad weather from a single image", in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2008.

[2] J.H. Lim, O. Tsimhoni and Y. Liu, "Investigation of driver performance with night vision and pedestrian detection systems—Part 1: Empirical study on visual clutter and glance behaviour", IEEE Trans. Intell. Transp. Syst., Vol. 11, pp. 670-677, 2010.

[3] S.G. Narasimhan and S.K. Nayar, "Vision and the atmosphere", International Journal of Computer Vision, Vol. 48, pp. 233-254, 2002.

[4] C.C. Chen, "Attenuation of Electromagnetic Radiation by Haze, Fog, Clouds, and Rain", A report prepared by UNITED STATES AIR FORCE PROJECT RAND, 1975.

[5] K. Rumar, "Infrared Night Vision Systems and Driver Needs", Warrendale, PA: SAE, 2003.

[6] J.M. Sullivan and M.J. Flannagan, "The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions", Accident Anal. Prev., Vol. 34, pp. 487–498, 2002.

[7] S.G. Narasimhan and S.K. Nayar, "Contrast restoration of weather degraded images", IEEE transactions on pattern analysis and machine intelligence, Vol. 25, No. 6, pp. 713-724, 2003.

[8] J. Davis and M. Keck, "A two-stage approach to person detection in thermal imagery", in Proc. Workshop on Applications of Computer Vision (WACV), 2005.

[9] Y. Wang, P.M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth and P. Ishwar, "CDnet 2014: An Expanded Change Detection Benchmark Dataset", in Proc. IEEE Workshop on Change Detection (CDW), pp. 387-394, 2014.

[10] S. Mahlke, D. Rösler, K. Seifert, J.F. Krems and M. Thüring, "Evaluation of six night vision enhancement systems: Qualitative and quantitative support for intelligent image processing", Hum. Factors, Vol. 49, pp. 518–531, 2007.

[11] O. Tsimhoni, J. Bärgman, M. Flannagan, "Pedestrian detection with near- and far-infrared night vision enhancement", Leukos, Vol. 4, pp. 113–128, 2007.

[12] S.G. Narasimhan, "Models and algorithms for vision through the atmosphere", Diss. Columbia University, 2003.

[13] R.C. Henry, S. Mahadev, S. Urquijo, D. Chitwood, "Color perception through atmospheric haze", JOSA A, Vol. 17, No. 5, pp. 831-835, 2000.

[14] J. Ge, Y. Luo, G. Tei, "Real-time pedestrian detection and tracking at night time for driver-assistance systems", IEEE Trans. Intell. Transp. Syst., Vol. 10, pp. 283–298, 2009.

[15] J. Han, A. Gaszczak, R. Maciol, S.E. Barnes and T.P. Breckon, "Human pose classification within the context of near-IR imagery tracking", in Proc. 9th SPIE 8901 Opt. Photon. Counterterrorism, Crime Fight. Def, pp. 1–11, 2013.

[16] M. Bertozzi, "IR pedestrian detection for advanced driver assistance systems", in Proc 25th DAGM Symp. Pattern Recognit, pp. 582–590, 2003.

[17] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, pp. –252, 1999.

[18] A.M. Elgammal, D. Harwood, L.S. Davis, "Non-parametric model for

background subtraction", in Proc. 6th European Conf. on Comput. Vis., pp. 751–767, 2000.
[19] M. Hofmann, P. Tiefenbacher and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 38–43, 2012.
[20] O. Barnich and M.V. Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences", IEEE Trans. Image Process., Vol. 20, pp. 1709–1724, 2011.
[21] P.L. St-Charles, G.A. Bilodeau and R. Bergevin, "SuBSENSE: A Universal Change Detection Method with Local Adaptive Sensitivity", IEEE Trans. on Image Processing, Vol. 24, pp. 359-373, 2015.
[22] P.L. St-Charles, G.A. Bilodeau, "Improving Background Subtraction using Local Binary Similarity Patterns", in Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV), 2014.
[23] A. Levin, Y. Weiss, F. Durand and W.T. Freeman, "Understanding Blind Deconvolution Algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 2354 – 2367, 2011.
[24] S. Li, "Markov Random Field Modeling in Computer Vision", Springer-Verlag, 1995.
[25] Z. Wu, N. Fuller, D. Theriault and M. Betke, "A thermal infrared video benchmark for visual analysis", in Proc. 10th IEEE Workshop on Perception Beyond the Visible Spectrum (PBVS), 2014.
[26] J. Portmann, S. Lynen, M. Chli and R. Siegwart R, "People detection and tracking from aerial thermal views", in Proc. IEEE International Conference on Robotics and Automation (ICRA), 2014.
[27] A. Berg, J. Ahlberg and M. Felsberg, "A thermal object tracking benchmark", in Proc. 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2015.
[28] P.L. St-Charles, G.A. Bilodeau and R. Bergevin, "Online Multimodal Video Registration Based on Shape Matching", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), 2015.
[29] C. Conaire, N. O'Connor, E. Cooke and A. Smeaton, "Comparison of fusion methods for thermo-visual surveillance tracking", in Proc. IEEE Conf. Information Fusion", 2006.
[30] J.W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery", Jour. Comp. Vis. and Im. Underst., Vol. 106, pp. 162 – 182, 2007.
[31] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu and A.M. ALópez, "Pedestrian Detection at Day/Nighttime with Visible and FIR Cameras: A Comparison", Sensors, Vol. 16, pi. E820, 2016.
[32] S. Hwang, J. Park, N. Kim, Y. Choi and I.S. Kweon, "Multispectral Pedestrian Detection: Benchmark Dataset and Baseline", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1037-1044, 2015.
[33] N. Goyette, P.M. Jodoin, F. Porikli, J. Konrad and P. Ishwar, "Changedetection.net: A new change detection benchmark dataset", in Proc. IEEE Workshop on Change Detection (CDW) at CVPR, Providence, RI, 2012.
[34] A. Singha and M.K. Bhowmik, "TU-VDN: Tripura University Video Dataset at Night Time in Degraded Atmospheric Outdoor Conditions for Moving Object Detection", in IEEE International Conference on Image Processing (ICIP), 2019 [Accepted].
[35] M. Heikkila and M. Pietikainen, "A texture-based method for modelling the background and detecting moving objects", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 28, pp. 657–662, 2006.
[36] G.A. Bilodeau, J.P. Jodoin and N. Saunier, "Change detection in feature space using local binary similarity patterns", in Proc. Conf. on Computer and Robot Vision (CRV), pp. 106–112, 2013.
[37] H. Wang and D. Suter, "A consensus-based method for tracking: Modelling background scenario and foreground appearance", Pattern Recognition, Vol. 40, pp. 1091–1105, 2007.
[38] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction", in Proc. IEEE Int. Conf. Pattern Recognit., pp. 28–31, 2004.
[39] D.S. Lee, "Effective gaussian mixture learning for video background Subtraction", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 27, pp. 827–832, 2005.
[40] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation", in Proc IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 302–309, 2004.
[41] M. Heikkila, M. Pietikainen and C. Schmid, "Description of interest regions with local binary patterns", Pattern Recognition, Vol. 42, pp. 425 – 436, 2009.
[42] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions", IEEE Transactions on Image Processing, Vol. 19, pp. 1635 – 1650, 2010.
[43] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, S.Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1301–1306, 2010.
[44] J. Baek, S. Hong, J. Kim and E. Kim, "Efficient Pedestrian Detection at Nighttime Using a Thermal Camera", Sensors, Vol. 17, 2017.
[45] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 21 – 26, 2012.
[46] K. Kim, T.H. Chalidabhongse, D. Harwood and L.S. Davis, "Background modeling and subtraction by codebook construction", in Proc. IEEE Int. Conf. Image Process, pp. 3061–3064, 2004.
[47] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction", Int. J. Electron. Commun., Vol. 64, pp. 739 – 747, 2010.
[48] N.M. Oliver, B. Rosario and A. Pentland, "A bayesian computer vision system formodeling human interactions", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 831–843, 2000.
[49] R. Radke, S. Andra, O. Al-Kofahi and B. Roysam, "Image change detection algorithms: a systematic survey", IEEE Trans. on Img. Processing, Vol. 14, pp. 294 –307, 2005.
[50] S. Herrero and J. Bescs, "Background subtraction techniques: Systematic evaluation and comparative analysis", In: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders, editors, Advanced Concepts for Intell. Vis. Syst., Springer Berlin Heidelberg, Vol. 5807, pp. 33–42, 2009.
[51] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications", Neural Networks, Vol. 13, pp. 411-430, 2000.
[52] E. Wigner, "On the distribution of the roots of certain symmetric matrices", The Annals of Mathematics, Vol. 67, pp. 325-327, 1958.
[53] A. Singha and M.K. Bhowmik, "Object Recognition based on Representative Score Features", in Proc. IEEE 18th International Conference on Advance Learning Technologies (ICALT), pp. 419-421, 2018.
[54] P.L. St-Charles, G.A. Bilodeau and R. Bergevin, "Universal Background Subtraction Using Word Consensus Models", IEEE Trans. on Image Processing, Vol. 25, pp. 4768-4781, 2016.
[55] W. Bin and D. Piotr, "A Fast Self-Tuning Background Subtraction Algorithm", in Proc. IEEE Int. Conf. on Comp. Vision and Pattrn. Recog., pp. 401-404, 2014.
[56] L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui and P. Sanchez, "Adaptive Multicue Background Subtraction for Robust Vehicle Counting and Classification", IEEE Trans. on Intelligent Transportation Systems, Vol. 13, pp. 527-540, 2012.
[57] F.C. Cheng, S.C. Huang and S.J. Ruan, "Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection", IEEE Trans. on Broadcastig, Vol. 57, pp. 794-801, 2011.
[58] W.H. Lee, "Foreground objects detection using multiple difference images," Jour. of Optical Engineering, Vol. 49, pp. 047201, 2010.
[59] Y. Goya, T. Chateau, L. Malaterre and L. Trassoudaine, "Vehicle trajectories evaluation by static video sensors", in Proc. 9th IEEE Int. Conf. on Intelligent Transportation Systems, pp. 864-869, 2006.
[60] S. Javed, A. Mahmood, T. Bouwmans and S.K. Jung, "Background–foreground modeling based on spatiotemporal sparse subspace clustering", IEEE Transactions on Image Processing, Vol. 26, pp. 5840-5854, 2017.
[61] H. Sajid and S.C.S. Cheung, "Universal multimode background subtraction", IEEE Transactions on Image Processing, Vol. 26, pp. 3249-3260, 2017.
[62] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian and Y. Yang, "Progressive Learning for Person Re-Identification with One Example", IEEE Transactions on Image Processing, Vol. 28, pp. 2872-2881, 2019.

Anu Singha received Master's Degree in Computer Applications, and Computer Science & Engineering from South Asian University (A SAARC University), New Delhi, and Tripura University (A Central University), Agartala, India in 2013 and 2015, respectively. From 2015, he is pursuing his Ph.D. in Computer Science & Engineering from Tripura University, India. Currently, he is also working as Junior Research Fellow (JRF) under a Defense Research and Development Organization (DRDO) funded project. His research interests include computer vision, object detection, face recognition, and deep learning etc.

Mrinal Kanti Bhowmik obtained his Bachelor of Engineering in Computer Science & Engineering from Tripura Engineering College in 2004 and Masters of Technology in Computer Science & Engineering from Tripura University (A Central University), India, in 2007. In 2014, he received Ph.D. (Engineering) degree from Jadavpur University, Kolkata, India. He has successfully completed two DeitY funded project, one DBT-Twinning funded project and one SAMEER funded project as Principal Investigator. Currently, he is the Principal Investigator of the two Govt. of India projects, one DRDO funded project and one ICMR funded project. From July 2010 onward, he is serving in the Department of Computer Science & Engineering at Tripura University as an Assistant Professor. His current research interests are in the field of computer vision, medical imaging, biometrics etc. He is also a Senior Member of the IEEE (USA).