Annotation and Benchmarking of a Video Dataset under Degraded Complex Atmospheric Conditions and Its Visibility Enhancement Analysis for Moving Object Detection

Sourav Dey Roy, *Student Member, IEEE*, and Mrinal Kanti Bhowmik, *Senior Member, IEEE*

***Abstract*— Detection of moving objects in outdoor environments is an extremely researched topic. However, studies on moving object detection in complex atmospheric/weather conditions are limited, primarily because of the absence of any relevant benchmark dataset. To address this disparity, we introduce a novel benchmark video dataset entitled "Extended Tripura University Video Dataset (E-TUVD)" which is a diverse dataset of complex atmospheric/weather conditions. Currently, E-TUVD is the largest video dataset for moving object detection under degraded atmospheric/weather conditions. The dataset comprises 147 video clips spanning 1-5 minutes in duration of each video clips. Because of the requirement of evaluating any object detection model, this study emphasizes on generation of ground-truth images of salient moving objects on E-TUVD. Using this dataset, we assessed the performance of several state-of-the-art algorithms, considering both the ability to detect moving objects and visibility enhancement under such complex conditions. The method with the best performance was used to investigate the effectiveness of visibility enhancement of atmospheric/weather degraded image sequences for accurate moving object detection. Results and analysis reveal that effective enhancement can significantly improve the ability of detection algorithms under degraded atmospheric/weather conditions to resemble the true properties of moving objects in terms of pixel oriented binary masks.**

***Index Terms*— Atmospheric/weather conditions, video dataset, meteorological information, ground-truth, moving object detection, visibility enhancement, performance evaluation.**

I. Introduction

MOVING object detection has been an active and mature research area in numerous computer vision applications because of the increasing demand of video surveillance for security applications. Fundamentally, it is often considered to be a pre-processing step and a low level task in computer vision applications, which is interconnected with high level inference tasks such as object localization, tracking, and classification. Its importance can be anticipated by visualizing the numerous articles published till date on this topic. Each moving object detection algorithm is designed to competently address the inherent real-world challenges of indoor/outdoor scenes, including illumination changes, dynamic backgrounds, ghosting artifacts, shadows, camouflage effects, etc. [1]–[3]. However, because outdoor scenes can be degraded by different complex atmospheric/weather conditions, moving object detection is more complicated under such conditions. Generally, the North-Eastern (NE) states, along with other states of India, share multiple international borders, because of which, security plays a vital role in such states. Under extreme atmospheric/weather conditions, outdoor scenes undergo from degradation, and suspicious intruders may not be detected by unaided human vision because of the high loss in contrast. Consequently, electronic surveillance plays an important role in detecting illegal threats to the state and for real-time detection of suspicious activities.

In spite of the blooming research on improving the impacts of predefined real-world challenges [2], [3], significant gaps exist in the existing solutions for detection of moving objects under atmospheric/weather degraded outdoor scenes. The rapid development of complex object detection algorithms originates from the available benchmark datasets that provides a balanced coverage of the challenges representative of the real-world [2]–[4]. Moreover, the availability of such dataset facilitates fair comparisons between state-of-the-art methods. Thus, the design of extensive datasets can provide a solid source for moving object detection and can consistently guide the development of this research field. In the last few decades, large datasets have been designed to meet the increasing demands for developing and benchmarking new models for moving object detection [4]–[38]. Each of these datasets is extensive in terms of amount or complexity. However, video datasets for moving object detection that can provide a balanced coverage of atmospheric/weather degraded outdoor scenes are still lacking.

Due to the remarkable effort required for designing an inclusive benchmark dataset to provide both pixel and location oriented ground-truth labels and a balanced coverage of

Manuscript received December 6, 2019; revised March 12, 2020; accepted April 23, 2020. Date of publication April 29, 2020; date of current version March 5, 2021. The work of Sourav Dey Roy was supported by the Council of Scientific and Industrial Research (CSIR), Government of India for providing the Senior Research Fellowship (SRF) under CSIR-SRF Fellowship Programme under Grant No. 09/714(0020)/2019-EMR-I, Dated: 01/04/2019. This article was recommended by Associate Editor B. Jeon. *(Corresponding author: Mrinal Kanti Bhowmik.)*

The authors are with the Department of Computer Science and Engineering, Tripura University (A Central University), Suryamaninagar 799022, India (e-mail: souravdeyroy49@gmail.com; mrinalkantibhowmik@tripurauniv.in; mkb_cse@yahoo.co.in).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2020.2991191

the representative atmospheric/weather challenges, previous attempts to objectively evaluate moving object detection methods in such complex situations have been restricted to limited partial assessments. Considering the importance of moving object detection for the computer vision and video processing communities, a video dataset under different atmospheric conditions with unambiguously defined moving objects must be created. Considering all these factors, the primary contributions of this study are summarized below:

1) We describe the designing issues and capturing protocol of a comprehensive real time benchmark video dataset of outdoor scenes entitled as "Extended Tripura University Video Dataset (E-TUVD)". The video dataset involves outdoor scenes degraded by various adverse atmospheric/weather conditions (such as fog, haze, dust, rain and poor illumination). The research community can utilize this dataset for testing and ranking of the existing and new algorithms for moving object detection in degraded atmospheric/weather conditions. To the best of our knowledge, E-TUVD is currently the largest moving object detection dataset for complex atmospheric/weather conditions.
2) Moreover, we provide a procedure for generating ground-truth images of suspected salient moving objects in terms of pixel oriented object masks and location oriented bounding boxes along with the object class in the maximum number of extracted frames of the created video dataset.
3) Based on this dataset, a comparison of thirty state-of-the-art methods (both classical and deep learning methods) for detection of moving objects in degraded atmospheric/weather conditions is provided based on pixel oriented simple matching coefficients which allow us to evaluate and develop new approaches on moving object detection in such adverse conditions.
4) We also provide a comparative analysis of twenty two state-of-the-art visibility enhancement methods (both classical and deep learning methods) based on no-reference image based quality assessment metrics. Furthermore, experiments were conducted to investigate the influence of visibility enhancement for accurate detection of moving objects in degraded complex atmospheric/weather conditions to identify the remaining challenges and provide the scope for future research.

Our goal is to offer a large and diverse benchmark dataset in adverse atmospheric/weather conditions to the research community to enable the design and evaluation of new moving object detection methods that can be readily used in the many real-time applications for such complex situations. A preliminary version of this paper has been previously published in [40]. Compared with the conference version, herein, we have extended the volume of the dataset and included some more atmospheric/weather challenges of real-world scenarios (i.e., rain and haze). In addition, in the present version, we provide a more comprehensive review of the existing dataset including experimental results for qualitative benchmarking of E-TUVD with respect to classical and deep learning based moving object detection and visibility enhancement methods. This article also aims to determine whether visibility enhancement of video sequences degraded by atmospheric/weather conditions can improve the ability of the detection method to accurately detect moving objects automatically. E-TUVD is a property of Tripura University, and details regarding this dataset are available in [41]. This dataset can be accessed for non-commercial use on request from [41].

The paper is organized as follows: In Section II, an overview of the existing datasets for moving object detection and the importance of our newly designed dataset in contrast to previous datasets are described. Section III, elaborated the designing issues and overall statistics of the created video dataset under different atmospheric/weather conditions. In Section IV, the procedure of generating ground-truth images (in terms of moving object masks and bounding boxes) of the salient moving objects in each of the extracted frames of E-TUVD is described. In Section V, thirty widely used state-of-the-art moving object detection techniques are analysed, and the experimental results of applying these methods to our dataset are detailed. Section VI compares the performance of twenty two state-of-the-art visibility enhancement techniques on E-TUVD for restoration of atmospheric/weather degraded extracted frames (i.e., due to fog, haze and dust). In Section VII, the effectiveness of visibility enhancement for the accurate detection of moving objects degraded by atmospheric/weather conditions is reported. Finally, Section VIII concludes the paper.

II. Related Work

A. Overview of Previous Datasets

Various datasets to evaluate the performance of moving object detection algorithms have been rapidly developed. Instead of elaborately discussing the large number of existing datasets on moving object detection, key features of each dataset based on a comparison of these datasets with our newly designed dataset (i.e. E-TUVD) are highlighted in TABLE I.

Importance of E-TUVD with respect to Previous Datasets:

Although the previous datasets reviewed in TABLE I have advanced the research in moving object detection, they have several drawbacks:

1) First, the datasets reviewed in TABLE I [4]–[38] are designed to provide predefined real-world challenges [2], [3], [39] (e.g., complex backgrounds, shadows, occlusion, intermittent object motion, activity of objects, etc.) in moving object detection or tracking algorithms. Although CD.Net 2014 [4], BMC 2012 [5], MarDCT [21], LASIESTA [25], and UWCD [38] consider some weather conditions (such as sunny, snowy, rain, cloudy and fog conditions), the number and variety of videos of adverse atmospheric/weather conditions are limited. Conversely, TU-VDN [33] contains video clips of adverse weather conditions (i.e., fog, dust, rain and low light), but this dataset specially focuses on key challenges like flat cluttered and dynamic backgrounds observed using static and

TABLE I

REVIEW AND COMPARISON ON EXISTING MOVING OBJECT DETECTION DATASETS USED IN RESEARCH WORKS WITH OUR DATASET

Frame/ Video Based Object Detection Dataset

Name of the Dataset	Key Characteristics	No. of Videos/ Scenes	Dataset Details				Total No. of Frames	Ground-truth Frames
			Environmental Conditions	Image Format	Dataset Type	Pixel Resolution		
CD.Net 2014 [4]	- Camera Jitter - Dynamic Background - Intermittent Object Motion - Shadow, Thermal and Night - Bad weather (Snow) and Low Frame Rate - Turbulence, Pan-tilt and Turbulence	53	Indoor and Outdoor	.jpg	V/T/C	320×240 to 720×480	159279	97334 (PBL)
BMC2012 [5]	- Complex Background - Climatic Conditions - Shadow and Crowded	20	Outdoor	.png	V/C	640×480	29980	15980 (PBL)
PETS2009 [6]	- Illumination Change and Occlusion - Crowded and Shadow - Intermittent Object Motion	8	Outdoor	.avi	V/C	720×576 to 768×576	NP	NP (BB)
I2R [7]	- Dynamic Background - Bootstrapping - Illumination Change	9	Indoor and Outdoor	.jpg	V/C	176×144	37958	37958 (PBL)
ETISEO [8]	- Crowded - Occlusion and Shadow - Illumination Change	118	Indoor and Outdoor	.mov	V/T/C	640×480	153243	153243 (BB, OC)
VSSN 2006 [9]	- Cluttered Background - Illumination Change - Bootstrapping	9	Indoor and Outdoor	.jpg	V/C	400×400	1000	1000 (PBL)
DAVIS [10]	- Cluttered Background - Motion Blur - Occlusion and Camera shake - Interacting objects	50	Outdoor	.jpg	V/C	1920×1080	3455	3455 (PBL)
Wallflower [11]	- Illumination change - Background motion - Camouflage Foreground Object - Bootstrapping	7	Indoor and Outdoor	.bmp	V/C	160×120	9917	7 (PBL) i.e. 1 Frame Per video
ViSal [12]	- Dynamic texture - Interacting objects - Crowded and Pose Variation	17	Indoor and Outdoor	.jpg	V/C/G	512×288	963	193 (PBL)
CAVIAR [13]	- Illumination Change - Crowded and Shadow - Intermittent Object Motion	80	Indoor	.mpg	V/C	384×288	152000	152000 (BB)
SegTrack [14]	- Complex deformation - Occlusion - Interacting objects	6	Outdoor	.png	V/C	320×240 to 414×352	244	244 (PBL)
SegTrack V2 [15]	- Appearance change - Complex deformation - Motion blur and Occlusion - Interacting objects	14	Outdoor	.png	V/C	259×327 to 640×360	976	976 (PBL)
FBMS [16]	- Occlusion - Illumination Change - Background motion	59	Indoor and Outdoor	.jpg	V/C	960×540	13860	720 (PBL)
VOS [17]	- Complexity of foreground - Background motion	200	Indoor and Outdoor	.mov	V/C	800×800	116103	7467 (PBL)
Fish4Knowledge [18]	- Blurred and Crowded - Complex background - Luminosity Change - Camouflage Foreground Object - Hybrid of all above	14	Underwater	.avi	V/C	320×240	NP	3500 (PBL) i.e. 250 frames per video
ViSOR [19]	- Illumination Change - Intermittent Object Motion - Shadow - Multiple Objects and Occlusion	623	Indoor and Outdoor	.avi	V/C	704×576	NP	1 frame per video (BB)
BEHAVE [20]	- Groups Interacting - Multiple Objects - Occlusion - Intermittent Object Motion	4	Outdoor	.wmv	V/C	640×480	83545	76800 (BB)
MarDCT [21]	- Complex Background - Blur and Haze - Occlusion	20	Outdoor	.m4v/.avi	V/T/C	704×576	NP	20 (PBL)
i-LIDS [22]	- Abandoned Baggage detection - Parked Vehicle detection - Doorway Surveillance - Sterile Zone monitoring	14	Indoor	.png	V/C	64×128	NP	Not labeled
SBM-RGBD [23]	- Bootstrapping - Camouflage and Illumination Change - Intermittent Object Motion - Out of sensors and Shadow	33	Indoor	.png	V/C/D	640×480	15000	1080 (PBL)
SBI2015 [24]	- Shadow and Occlusion - Camera Jitter - Dynamic Background	14	Indoor and Outdoor	.png	V/C	200×164 to 2272×1704	8548	14 (BG)
LASIESTA [25]	- Camouflage and Occlusions - Illumination Changes and Bootstrapping - Camera motion and Camera Jitter - Weather (Sunny; Snow; Rain; Cloudy)	48	Indoor and Outdoor	.bmp	V/C	352×288	18725	18725 (PBL)
E-TUVD [Our Dataset] (Still Growing)	- Weather Challenges (Fog; Haze; Dust; Rain; Poor Illumination; Clear Day) - Background Challenges - Other Real World Challenges	147	Outdoor	.mov	V/C	1920×1080	793800 (Approx.)	1 frame per 5 frames (PBL, BB, OC)

PBL- Pixel Based Labeling; BB- Bounding Box; OC- Object Class; BG- Background; SM- Shadow Mask; V- Visual; T-Thermal; C-Color; D- Depth; G- Gray; NP- Not Provided

TABLE I

(Continued.) REVIEW AND COMPARISON ON EXISTING MOVING OBJECT DETECTION DATASETS USED IN RESEARCH WORKS WITH OUR DATASET

Frame/ Video Based Object Detection Dataset

Name of the Dataset	Key Characteristics	No. of Videos/ Scenes	Dataset Details				Total No. of Frames	Ground-truth Frames
			Environmental Condition	Image Format	Dataset Type	Pixel Resolution		
UCSD [26]	- Dynamic Background - Motion Blur	18	Outdoor	.jpg	V/G	232×152 to 468×348	1927	886 (PBL)
ATON [27]	- Illumination Changes and Blur - Shadow and Multiple Objects	5	Indoor and Outdoor	.avi	V/C	320×240	NP	113 (PBL)
LIMU [28]	- Illumination Changes and Blur - Shadow and Crowded - Occlusion	8	Indoor and Outdoor	.jpg	V/C	320×240	37801	2257 (PBL)
SZTAKI [29]	- Illumination Changes - Shadow and Multiple Objects	5	Indoor and Outdoor	.jpg	V/C	320×240	861	861 (PBL, SM)
VIVID [30]	- Camouflage and Shadow - Occlusion and Crowded - Intermittent Object Motion	9	Outdoor	.jpg	V/T/C	640×480	16274	1633 (PBL, BB)
CAMO-UOW [31]	- Camouflage and Shadow - Multiple Objects	10	Indoor and Outdoor	.jpg	V/C/G	1600×1200 to 1920×1080	3517	3517 (PBL)
GTFD [32]	- Intermittent Object Motion - Shadow and Thermal Crossover - Cluttered Background and Dynamic Scene	25	Indoor and Outdoor	.png	V/T/C	384×228	1067	1067 (PBL)
TU-VDN [33]	- Cluttered and Dynamic Background - Weather Challenges (Fog; Dust; Rain; Low Light)	60	Outdoor	.mov	T/G	640×480	138230	22030 (PBL)
REMOTE SCENE IR [34]	- Camouflage and Video Noise - Dynamic Background and Camera Jitter - Ghost Artifacts	12	Outdoor	.bmp	T/G	640×512	1263	1263 (PBL)
GSM [35]	- Dynamic Background - Illumination Changes and Bootstrapping - Camouflage and Shadow	7	Indoor	.png	V/C/D	544×135 to 552×136	3361	87 (PBL)
SBMnet [36]	- Intermittent Object Motion - Cluttered Background and Camera Jitter - Illumination Changes - Dynamic Background	79	Indoor and Outdoor	.jpg	V/C	360×240 to 800×600	73357	19 (BG)
KINDERGARTEN [37]	- Shadows and Illumination Changes - Intermittent Object Motion - Camouflage and Noise	100	Indoor and Outdoor	.avi	V/C	352×288	NP	NP
UWCD [38]	- Illumination Changes - Dynamic Background and Shadow - Camouflage and Bad weather (Snow)	5	Underwater	.png	V/C	1920×1080	5507	500 (PBL)
E-TUVD [Our Dataset] (Still Growing)	- Weather Challenges (Fog; Haze; Dust; Rain; Poor Illumination; Clear Day) - Background Challenges - Other Real World Challenges	147	Outdoor	.mov	V/C	1920×1080	793800 (Approx.)	1 frame per 5 frames (PBL, BB, OC)

PBL- Pixel Based Labeling; BB- Bounding Box; OC- Object Class; BG- Background; SM- Shadow Mask; V- Visual; T-Thermal; C-Color; D- Depth; G- Gray; NP- Not Provided

motion cameras. This is because all video clips of TU-VDN are recorded with a far-wave infrared sensor at night time. In addition, because wavelengths in the far-infrared region exceed those in other infrared wave bands, impact of atmospheric particles on far-infrared waves is relatively insignificant. Therefore, such videos may not be pragmatic in real-world outdoor scenarios and may cause over-fitting of the moving object detection methods because moving objects detection in outdoor environments are mainly susceptible to atmospheric/weather effects. In comparison, our newly designed video dataset specially focuses on various adverse atmospheric/ weather challenges, including the above mentioned predefined challenges [2], [3], [39] of moving object detection for the evaluation of standard baseline models under such complex situations.

2) Second, ground-truth annotation of foreground moving objects in previous datasets is provided either in terms of bounding boxes or object masks. Thus these datasets focus on either tracking or detection algorithms. Although foreground moving objects in some datasets are annotated in the form of binary object masks, they are often manually annotated by one subject, resulting in strong subjective bias with respect to performance evaluation of the detection models. In our newly designed dataset, we first localized the salient moving objects using rectangular bounding boxes, and then, labeled pixel wise annotation of the moving objects present in the bounding boxes using multiple annotators such that the resultant ground-truth image sequences become the most probable joint agreement of all annotators.
3) Third, these datasets are less significant than the proposed dataset for traditional Convolution Neural Networks (CNNs) because the number of frames and annotated frames in most of the previous dataset mentioned in TABLE I are not sufficiently large for training traditional deep learning models. To cope up with such problems, manual labeling (bounding boxes and binary object masks) of the maximum number of frames in each of the video clips of E-TUVD dataset are provided with the E-TUVD.

In summary, the existing datasets are not sufficient for benchmarking the performance of traditional moving object detection methods in degraded complex atmospheric/weather conditions because of the less variety of available videos in such degraded conditions as well as the ambiguous annotation

of salient moving objects. For advance this area, E-TUVD provides a balanced coverage of real-world outdoor scenarios in degraded atmospheric/weather conditions and contains ground-truth images of salient moving objects, which are unambiguously defined and annotated.

B. Contributive Features of E-TUVD

The main contributive features of the designed Extended Tripura University Video Dataset (E-TUVD) are:

1) The dataset comprises video sequences of moving objects (especially vehicles, pedestrians and animals) under various atmospheric/weather degraded challenges captured from different security and surveillance zones of Tripura.
2) The dataset contains 147 videos including 29 videos in clear day conditions, 22 videos in poor illumination conditions, 34 videos in fog conditions, 12 videos in haze conditions, 23 videos in dust conditions, and 27 videos in rain conditions captured under static and dynamic background conditions.
3) Along with atmospheric/weather challenges, the dataset also contains various representative challenges of real-world moving object detection i.e., intermittent object motion, camera jitter, shadows and overlap of two moving objects.
4) Each frame of the E-TUVD contains multiple types of moving objects, e.g., pedestrians, bikers, bicyclists, cars, buses, trucks, autos, rickshaws, and animals (i.e., dogs, cows, and cats). In addition, the scenes were captured mostly in urban areas, which are subjected to larger surface variations because of the presence of objects such as trees, houses, warehouses, office buildings, streets, and residents.
5) For each video clip in the dataset, meteorological data regarding atmospheric/weather information (i.e., dew point, temperature, humidity, etc.) on the capturing day were obtained from meteorological department of Tripura State and are also provided.
6) For all the captured video sequences, ground-truth images of salient moving objects are annotated in terms of binary object masks and object fixations (i.e., location oriented bounding boxes) to alleviate ambiguities in defining and annotating salient moving objects and are provided along with this dataset.

All these features of E-TUVD reflect its significance in the moving object detection from various real-world atmospheric/weather degraded outdoor scenes.

III. Designing Issues and Overall Statistics of E-TUVD in Degraded Atmospheric/Weather Conditions

Generally, outdoor images often suffer from low contrast and poor visibility information because the light reflected from the object of interest (in our case, the moving objects) is attenuated in air and further blends with the atmospheric light scattered by some aerosols or water-droplets before it reaches the camera. Inclement weather has remained a challenge for many computer vision applications because the performance of imaging devices drastically reduces under such conditions. In this section, we elaborate the capturing conditions, acquisition set-up, overall statistics, and naming conventions of the E-TUVD under atmospheric or weather degraded conditions.

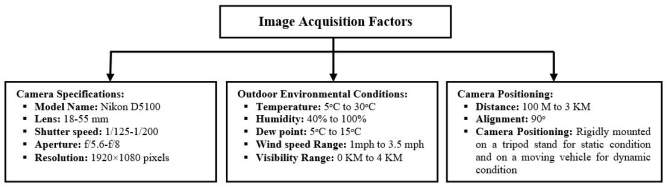

Fig. 1. Acquisition factors of E-TUVD [40].

A. Dataset Capturing Conditions

Images in an outdoor environment are mainly susceptible to two factors: Weather and Illumination. The weather effect is due to the presence of fog or other associated particles suspended in air that scatter light in the atmosphere [42]. In contrast, the illumination effect in an outdoor environment mainly occurs because of variations in the intensity of sunlight at different times of the day, i.e., specular reflection [43]. Such conditions alter the key characteristics (i.e., intensity, color, polarization and coherence) of sunlight because of scattering of light by atmospheric particles. The atmospheric or weather conditions considered in our E-TUVD dataset are: Fog Conditions, Haze Conditions, Dust Conditions, Rain Conditions and Poor Illumination Conditions. In addition to the above mentioned five atmospheric or weather conditions, E-TUVD also contains video clips of clear days where horizontal sunlight casts long shadows and gives subjects a warm glow with high contrast scenes. Considerable effort was invested for measuring the physical properties of these atmospheric/weather conditions (i.e., types and sizes of particles involved and their concentrations in space) [43].

B. Camera Set-up for Dataset Acquisition

To maintain uniformity and collect valuable information for our research work, the effects of various factors considered during data acquisition to facilitate the observation of atmospheric/weather on scene appearances are shown in Fig. 1.

The acquisition of the video dataset under the atmospheric/weather conditions involved of three components. The first component is the camera specification. Herein, video clips for the E-TUVD were captured using a Nikon D5100 camera (Thailand) which had a focal length of 18-55 mm with a horizontal field-of-view, zoom ratio of 3.00x and aperture range of f/5.6-f/8. The video resolution was 1920 $\times$ 1080 pixels and the video images were not stabilized during acquisition. The second component is the outdoor environment condition. Depending on different atmospheric conditions, the outdoor temperature during data acquisition varied from 5°C to 30°C, humidity was between 40% and 100%, the dew point ranged from 5°C to 15°C,

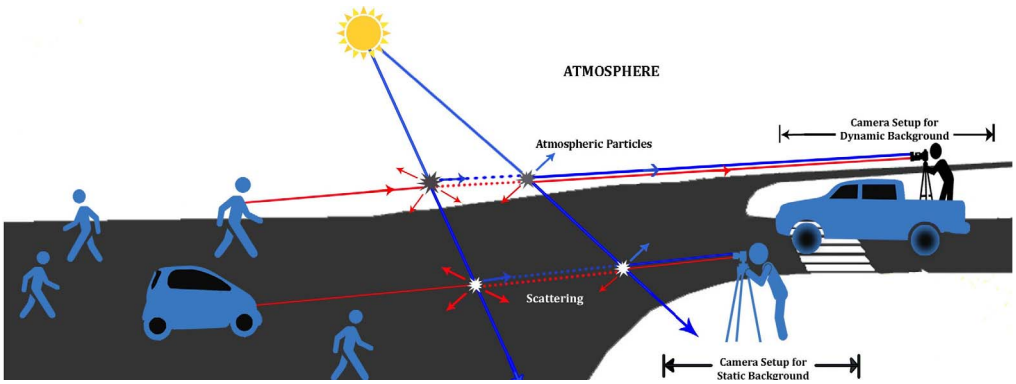

Fig. 2. Overall acquisition setup for dataset capturing of E-TUVD under degraded atmospheric/weather conditions.

Fig. 3. Sample image frames of E-TUVD under different Atmospheric/Weather conditions: row (1) Fog conditions; row (2) Haze conditions; row (3) Dust conditions; row (4) Rain conditions; row (5) Poor Illumination conditions; row (6) Clear Day conditions.

the wind speed varied from 1 mph to 3.5 mph, and the visibility remained approximately between 0 KM to 4 KM. The third component is positioning the camera with respect to the targeted objects. During data acquisition, the alignment of the camera (i.e., line of sight) with respect to the moving objects was horizontal over path lengths of several kilometers (i.e., approximately 3 KM). The camera was positioned on a tripod stand with vertical height adjustment and an average elevation of 5 feet from the ground level. The video clips of the E-TUVD were captured under two background conditions (i.e., static and dynamic background). For capturing the video clips in a static background condition, the camera was kept fixed with respect to the moving objects by mounting the camera on a tripod stand (in this case, the background is static with respect to the moving objects). Conversely, to capture video clips under dynamic background conditions, the video clips were captured by mounting the camera on a moving vehicle, and the speed of the vehicle was maintained between 20 and 30 KM/H. In this case, both the objects and the background are moved simultaneously. The jagged and diagrammatic representation of the complete set-up for data acquisition is displayed in Fig. 2.

The video clips of the E-TUVD were collected from different locations of Tripura, which is a North Eastern State of India, and were captured from the Tripura University Campus and different crossings of Agartala city as per availability of data under the aforementioned degraded atmospheric/weather conditions throughout the year. Some of the sample frames of the E-TUVD depicting different adverse atmospheric/weather conditions are shown in Fig. 3. The image frames are from the beginning, middle and end of the video sequences for visualization.

C. Dataset Statistics

According to the above mentioned acquisition factors, currently, E-TUVD contains 147 video clips, depicting outdoor scenes with moving objects that were captured under different scenarios and include various challenges under different

TABLE II
DISTRIBUTION OF E-TUVD IN DIFFERENT ATMOSPHERIC/WEATHER CONDITIONS

Image Type	Camera Model	Background Condition	Atmospheric/ Weather Conditions						Total Videos
			Fog Conditions	Haze Conditions	Dust Conditions	Rain Conditions	Poor Illumination Conditions	Clear Day Conditions	
Visual	Nikon D5100	Static	24	7	14	15	15	18	93
		Dynamic	10	5	9	12	7	11	54
Total Number of Videos			**34**	**12**	**23**	**27**	**22**	**29**	**147**

**As in, May 2019 (Dataset is still growing)

atmospheric/weather conditions (as shown in Fig. 3). Each video clip was 1 to 5 minutes long with a frame rate of 30 fps (Frame per Second). The overall statistics of the created dataset are shown in TABLE II. The key challenges of the designed video dataset are broadly categorized as follows:

- **Background Challenges:** The E-TUVD video clips depict outdoor scenes either under static or dynamic background conditions. Handling background dynamics in which the background is continuously moving with respect to the moving objects is a challenging task. The dataset contains 93 video clips under the static background conditions and 54 video clips under dynamic background conditions.
- **Atmospheric/Weather Challenges:** The dataset includes urban scenes with buildings, trees, sky, vehicles, and pedestrians with ranges from approximately 100 meters to 3 KM. The dataset contains 22 video clips under the poor illumination conditions, 34 video clips in fog conditions, 12 video clips in haze conditions, 23 video clips in dust conditions, and 27 video clips in rain conditions, as shown in Fig. 3. Moreover, the dataset also contains 29 video clips on a clear day to facilitate the comparison of complex moving object detection models under both clear conditions and atmospherically degraded conditions.
- **Other Challenges:** In addition to these two major challenges, the dataset also contains scenes with multiple moving objects in a single frame, overlapping or occluding of the moving objects in a single frame, shadow, and intermittent motions of objects. Although camouflage or poorly textured and camera jitter images are some of the major issues considered for moving object detection, we did not consider them in the categorical challenges of the E-TUVD because almost all the real-world video sequences contain some level of camouflage and camera jitter effect.

Each of these subsets was very challenging and could be used to test benchmark moving object detection/tracking algorithms under realistic scenarios.

D. Naming Convention

Each category of video clips in the E-TUVD was named to ensure that the dataset could be understood during analysis. Different codes were assigned to the different atmospheric condition, different dates on which data were captured, and different types of background. Using these codes, the dataset were named in the format Capturing-Day_Atmospheric-Condition_ Background-Type_Video-ID.mov. All the assigned codes for each component of the name are illustrated in Table III. Using the above naming convention, every video clip in the dataset acquired a distinct identity. Based on the codes provided in Table III, the image name "D1_F_S_01.mov" indicates that the video is Video_ID 01, which is a static background video captured under fog conditions on the first day.

TABLE III
CODES USED FOR NAMING E-TUVD

Atmospheric Condition		Background Condition		Capturing Day	
Condition	Codes	Type	Codes	Day	Codes
Fog	F			Day1	D1
Haze	H			Day2	D2
Dust	D	Static	S	Day3	D3
Rain	R	Dynamic	D	Day4	D4
Poor Illumination	PI			---	---
Clear Day	CD			Dayn	Dn

E. Meteorological Information

Analyzing outdoor scenes under extreme atmospheric/ weather degraded conditions requires metrological data from the capturing location or in the area adjacent to the capturing location. According to the World Meteorological Organization (WMO) network, atmospheric/weather phenomena are depicted and measured using different variables of the earth's atmosphere; i.e., Temperature (°C), Humidity (%), Dew Point (°C), Wind Speed (KM/H), Wind Direction, Atmospheric Pressure (mbar) and Visibility (KM). Each video clip of the E-TUVD contains useful ground-truth meteorological information related to these atmospheric variables/ parameters obtained from the Regional Meteorological Department of Tripura, which is provided with the dataset.

IV. GROUND-TRUTH GENERATION OF SALIENT MOVING OBJECTS ON THE E-TUVD DATASET

By generating ground-truth images of salient moving objects, the effectiveness of moving object detection/tracking algorithms can be verified. The most common and efficient method for generating ground-truth images is using manual annotation. However, manual annotation of accurate ground-truth data that contain the exact area of salient moving objects often results in uncertainty and strongly subjective bias. Thus, generating uncontroversial ground-truth images that provides the maximum information regarding salient moving objects in each frame is one of the most tedious and challenging tasks. In our study, we collected two types of ground-truth data, moving object fixation using rectangular bounding boxes

and pixel oriented binary moving object masks to mitigate the indistinctness in defining and annotating salient moving objects.

A. Rectangular Bounding Box Based Moving Object Fixation

Currently, the dataset contains 147 video clips 1-5 minutes each (30 fps). Therefore, it is very difficult and time consuming for a single person to accurately annotate ground-truth images of salient moving objects for the captured videos. Moreover, manual annotation of salient moving objects in videos of complex outdoor scenes is a difficult and challenging task. This is because several candidate objects may exist and different annotators may be biased in determining which objects are "salient" or "moving". Therefore, for consistent annotation of the E-TUVD, a more general definition of "salient moving objects" in a video is essential to guide the manual annotation processes. First, to address the problem of "what are the most likely salient objects present on our E-TUVD dataset", we visually analyzed all video clips of the E-TUVD and predefined the classes of salient objects to be annotated. They are – Pedestrian, Biker, Bicyclist, Car, Bus, Truck, Auto, Rickshaw and Animal (i.e., Dog, Cow and Cat). Second, it is particularly difficult to reliably identify the actual number of moving objects in a multi object frame because all objects may not move simultaneously. In addition, the salient moving objects may become immobile in the middle of a video clip (i.e., intermittent object motion). To approximate the salient objects that are "moving" (i.e., salient moving objects) well in the video frames of the E-TUVD, the background model was estimated using a previously developed model [44] and compared with current images. This eased the task of annotators to identify the regions of interest (i.e., moving objects) in a frame. Third, assigning the classes of salient moving objects in a multi object frame is difficult because in some situations, a small part of a salient moving object may be visible or may be occluded by some other candidate objects. Under such circumstances, annotators were instructed to carefully observe the previous or next successive frames and assign the object to one of the aforementioned predefined classes. Depending on this information, manual fixation of the salient moving objects in the form of a rectangular box was performed, and its identity and detailed spatial and temporal information were provided.

To implement the protocol, five members of the research laboratory were selected and trained to maintain consistency in ground-truth annotation. They were provided a set of annotation guidelines, consisting of what to annotate, what are the class labels and how to handle occlusions. The six categories of video clips from E-TUVD (i.e., fog, haze, dust, rain, poor illumination and clear day) were equally distributed among these five individuals. Each annotator was asked to free-view all the extracted frames of the video clips provided to them and create a rectangular boxes for each salient moving objects present in one frame per five frames using the LabelImg graphical image annotation tool [45]. This graphical user interface based tool is freely available [45]. The rectangular bounding boxes were created by fixing the coordinates of two points (i.e., recommended for PascalVOC) of each salient moving object in one frame per five frames and defining the class of these bounding boxes. These two points are: the upper most left corners (xmin, ymin) and the lower most right corners (xmax, ymax). Along with bounding box information (i.e. the coordinates of the two points) outlining the salient moving objects, temporal information related to Frame_ID, Number_of_Objects, and Object_Class as present in the corresponding frames for each video clip of E-TUVD was also maintained in an .xml file and provided with the dataset. During the entire process, annotators were periodically observed to ensure they followed the guidelines.

B. Generation of Binary Moving Object Masks

To generate accurate and uncontroversial ground-truth images of the salient moving objects in the form of binary moving object masks, the most suitable method is to combine the manually generated binary moving object masks of all subjects so that the resultant ground-truth object mask is most likely to each of the annotator segmentation. Considering this aspect, one of the most well-known semi-automated tool, Tool for Semiautomatic Labeling (TSLAB) [46], was used to generate the binary moving object masks. This is a freely available graphical user interface based tool [47]. To implement the protocol, five members of the research laboratory who did not participate in annotating the bounding boxes were selected and trained for binary object mask generation. Similar to the previous subsection, the six categories of video clips from the E-TUVD were equally distributed among these five annotators. Along with the video clips, bounding box related information was also provided to avoid the problem of determining the salient moving objects. Using TSLAB [47], annotators were asked to manually draw the contours of each salient moving object present in the respective frame in the VMO (Visible Moving Object) layer selection mode. Consequently, the area inside the contours was considered as moving (i.e., each of the salient moving objects was annotated as a whole based on a semantic concept). Thus, in each video clip of the E-TUVD, ground-truth images of one frame per five frames were produced in a .bmp format with the following two labels: Static Pixels: Assigned the binary values of 0 and Moving Pixels: Assigned binary values of 1.

After generating the ground-truth images (i.e., the binary moving object masks), the video clips of the E-TUVD along with the corresponding bounding box information were exchanged among the workers five times to generate five reference ground-truth images corresponding to each frame (one frame per five frames) of the E-TUVD video clips. Once the five software based reference ground-truth images were obtained, the maximum voting policy was used to combine the results of these five ground-truth images. In this method, a threshold value T (in our case T=3) was set depending on the number of reference ground-truth images, and a ground-truth image was built as the maximum likelihood estimation. This can be mathematically represented as follows:

$$\mathrm{GT}(i,j) = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} \mathrm{RGT}_k(i,j) \geq \mathrm{T} \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

where $RGT_k(i,j)$ is the k^{th} reference ground-truth image. The resultant maximum voting policy based binary moving object masks of one frame per five frames for each video clips of E-TUVD are provided in the .bmp format with the dataset.

V. Detection of Moving Objects Using the State-of-the-Art Methods in Degraded Atmospheric/Weather Outdoor Conditions on E-TUVD

A. State-of-the-Art Methods Tested

Detection of moving objects has been extensively studied since the 1990's; consequently, numerous new ideas have been proposed for high-quality detections of complex scenarios with the lowest misdetection rate. The simplest object detection strategy is to segment the moving regions of interest from the static background. We reviewed previous publications [48]–[54] that presented survey on primitive approaches/categories for moving object detection from videos and have reported the general concepts of various representative methods for each category. Although many detection techniques are used in computer vision, not all work well for atmospheric/weather degraded complex situations. Most of these techniques are image and application dependent, and their results are highly dependent on image characteristics. From the literature, a total of 30 state-of-the-art methods (i.e., 26 classical methods and 4 deep learning methods) were implemented and tested on the E-TUVD to obtain the most efficient technique(s) for the detection of moving objects in degraded atmospheric/weather conditions. Some of these methods indicated by an asterisk (∗) in TABLE IV were implemented in BGSlibrary [54]. The remaining techniques are implemented by own or as per the codes made available by the corresponding authors. For each method, only one set of parameters was used for all the videos. These parameters were selected based on author recommendations, or, when not available, were adjusted to enhance the overall results. A brief overview of each technique along with the associated parameters and their settings is provided in TABLE IV.

To better visualize the categorization results, Fig. 4 depicts the displayed results of thirty state-of-the-art detection methods on E-TUVD under various atmospheric/weather conditions.

B. Evaluation Metrics

The performance of moving object detection methods proposed in the literature has been evaluated either based on visual or objective inspection of the detection results. Because none of the existing state-of-the-art methods are relevant to all the situations of the E-TUVD, objective evaluation of the detection algorithms is essential. Similar to many pixel based moving object detection algorithms, we also adopted the three most widely used supervised or reference image based evaluation metrics (i.e., pixel oriented simple matching coefficient based qualitative measurement). These evaluation metrics are [69], [70]: F-Measure (F-Mes), Percentage of

TABLE IV

Brief Summary of the Moving Object Detection Techniques Used in the E-TUVD Dataset With the Associated Parameters

Method ID	Method Name		Approach/ Technique	IC	Parameters Setting
ViBe*	Visual Background Extractor	[44]	BM/ PLOD	P	R- 20; S_F- 16; N- 20; N_C- 2
ISBM	Illumination Sensitive Background Modeling	[55]	BM/ PLOD	M	T_1- 0.05, T_2- 20
MTD	Multiple Temporal Difference	[56]	FD/ PLOD	M	T- 20
PBAS*	Pixel Based Adaptive Segmenter	[57]	BM/ PLOD	P	N- 35; Min- 2; $R_{I/D}$- 0.05; R_L- 18; R_S- 5; T_D- 0.05; T_I- 1; T_L- 2; T_U- 200
GMM-V1	Stauffer- Gaussian Mixture Model	[58]	BM/ PLOD	M	T- 9; LR- 0:05; η- 3
BM	Bayesian Modeling	[59]	BM/ PLOD	M	LF- 20; LR- 0.05; γ- 0.1
XCS-LBP	eXtended Center-Symmetric Local Binary Pattern	[60]	BM/ PLOD	M	X_R- 1; Y_R- 1; T_I- 2; T_L- 2; B_L- 1; N_P- 8;
DECOLOR	DEtecting Contiguous Outliers in the LOw-rank Representation	[61]	BM/ PLOD	M	λ- 1; C_P- 1e-4; N_I- 20; N_O- 50
GMM-V2	Zivkovic- Gaussian Mixture Model	[62]	BM/ PLOD	M	T- 20; LR- 0:01
EB	Eigen Background	[63]	BM/ PLOD	M	T- 255; HS- 10; ED- 10
SC-RPCA	Segmentation Constrained Robust PCA	[64]	BM/ PLOD	M	μ- 10^{-6}; p- 1.1
IMTSL	Incremental and Multi-feature Tensor Subspace Learning	[65]	BM/ PLOD	P	W_S- 25; r1- 1; r2- 8; r3- 2; $t_1/t_2/t_3$- 0.01; $w_1/w_2/w_3/w_6/w_7/w_8$- 0.125; w4- 0.225; w5- 0.025
LOBSTER*	Local Binary Similarity Segmenter	[66]	BM/ PLOD	P	T_{desc}- 12; T_{int}- 90; N_{LSBP}- 35; min- 2; φ- 16; T- 0:5; T_r- 0:365; T_d- 30
MBS	Multimode Background Subtraction	[67]	BM/ PLOD	P	corr_th-0.99; prob_th-0.75; M-300; col_th-3; ch_th- 100; T- 0.15; m- 0.5
SSS-RPCA	Spatiotemporal Structured-Sparse RPCA	[68]	BM/ PLOD	M	$γ_1$/ $γ_2$-0.08; ϵ_1- 10; r_p-8; K_{TW}-200; S-250
PAWCS*	Pixel based Adaptive Word Consensus Segmenter	[69]	BM/ PLOD	P	t_0- 1000; N_W- 50; R_C- 20; R_d- 2; α- 0.01; $λ_T$- 0.5; $λ_R$- 0.01
SUBSENSE*	Self-Balanced Sensitivity Segmenter	[70]	BM/ PLOD	P	T_r- 0.333; R_c= 3; R_d- 30; N- 50; #min- 2; α^{LT} - 100
CB*	Codebook	[71]	BM/ PLOD	P	N_W- 6.5 per pixel; TF- 40; T_M- 20; ϵ- 20
VU-METER*	Non Parametric Probabilistic Model	[72]	BM/ PLOD	P	T- 0:03; LR- 0:995; B_S- 8
KDE*	Kernel Density Estimation	[73]	BM/ PLOD	M	Non Parametric
MULTICUE*	Adaptive multicue strategy	[74]	BM/ PLOD	P	T1- 5; T2- 0.2
SOBS	Self-organization through artificial neural networks	[75]	BM/ PLOD	P	LR- 180; LR_T- 255; p-100; p_T - 240; TS- 40
SC-SOBS	Spatially Coherent Self-Organizing Background Subtraction	[76]	BM/ PLOD	P	LR- 180; LR_T- 255; p-100; p_T - 240; TS- 40
KNN*	K-Nearest Neighbor	[77]	BM/ PLOD	P	T- 20; LR- 0:01; η- 3
LBP-MRF*	Local Binary Pattern and Markov Random field	[78]	BM/ PLOD	P	R- 2; K- 3; T_1- 0.95; T_2- 0.75; LR_1- 0.01; LR_2- 0.01; γ- 8
WMV*	Weighted Moving Variance	[79]	BM/ PLOD	P	T- 15
IDLM	Interactive Deep Learning Method	[80]	DL/ PLOD	M	NR
MFCN	Multiscale Fully Convolutional Network	[81]	DL/ PLOD	M	NR
MSFgNet	Motion Saliency Foreground Network	[82]	DL/ PLOD	P	NR
ED-ResNet18	ResNet-18 via Encoder Decoder Structure	[83]	DL/ PLOD	P	NR

IC- Implemented Code; BM- Background Modeling; DL- Deep Learning; PLOD- Pixel Level Object Detection; P- Python; M- MATLAB; R- Radius of a Sphere; S_F- Sampling Factor; N_C- Number of close pixel samples; T/T1/T2/ T_M- User Settable Thresholds; N- Number of Background Components; Min- Number of components closer to Decision Threshold; $R_{I/D}$- Rate at which Decision Threshold is Regulated; ϵ- Detection Threshold; T_r- Relative Threshold; #min- Minimum Number of Matches for Background Classification; α^{LT}- Number of Samples for Moving Average; R_L- Lower Bound of Decision Threshold; R_S- Scaling Factor of Decision Threshold; T_D/ T_I- Decreasing/ Increasing Rate at which Probability of background is Updated; T_L/ T_U- Lower/ Upper Bound of Learning Parameter; TL- Time Length; BL- Border Length; N_P- Neighboring Points; λ- Regularization Parameter; C_P- Convergence Precision; N_I/ N_O- Maximum Number of Inner and Outer Iterations; W_S- Block Size; $γ_1$/ $γ_2$- Parameter Constant; ϵ_1- Nearest Neighbors for Constructing the Graphs; rp- Number of Representation; ; K_{TW}-Non Overlapping Temporal Windows; S/ M- Number of Super Pixels; corr_th-Correlation Threshold; prob_th- Probability Threshold; col_th- Color Threshold; ch_th- Threshold of Color Channel; m- Detection Threshold; t_0- Word weight Offset Value; N_W- Maximum number of words per dictionary; R_C- Color Distance Threshold; R_d- LSBP Distance Threshold; α- Adaptation Rate; $λ_T$- Updation rate Change Factor; $λ_R$- Distance Threshold Change factor; X_R/ Y_R/ T_I- Radii Parameter along X, Y and T axis; TI- Time Interval; η- Number of Gaussians; HS- History Size; ED- Embedded Dimension; μ- Penalty Parameter; p- Convergence Speed Controller; LF- Number of Learning Frames; LR/ LR_1/ LR_2- Learning Rate; LR_T- Learning Rate of Training Frames; TS- Number of Training Steps; p_{T^-} Sensitivity; p_{T^+} Sensitivity of Training Frames; B_S- Bin Size; R- Radius of a circle from which LBP histogram is computed; K- Number of pieces of Histogram; γ- Constant; T_{desc}/ T_{int}- Threshold to determine if an input pixel matches the model based on the Hamming/ L1 distance; N_{LSBP}- Number of LSBP Descriptors; φ- Sampling Factor; T_r/ T_d- Relative/ Absolute LSBP Descriptor Threshold; r1/r2/r3- Desired Ranks; $t_1/t_2/t_3$- Corresponding Thresholds of Ranks; W_1 to W_8- Set of Weights; NR- Not Required for Model Testing.

Fig. 4. Typical segmentation results for various atmospheric conditions on the E-TUVD: **Left side of the figure:** row (1) shows input Frames; row (2) shows Ground-truth; row (3) shows ViBe results [44]; row (4) shows ISBM results [55]; row (5) shows MTD results [56]; row (6) shows PBAS results [57]; row (7) shows GMM-V1 results [58]; row (8) shows BM results [59]; row (9) shows XCS-LBP results [60]; row (10) shows DECOLOR results [61]; row (11) shows GMM-V2 results [62]; row (12) shows EB results [63]; row (13) shows SC-RPCA results [64]; row (14) shows IMTSL results [65]; row (15) shows LOBSTER results [66]; row (16) shows MBS results [67]; row (17) shows SSS-RPCA results [68]; **Right side of the figure:** row (1) shows Input Frames; row (2) shows Ground-truth; row (3) shows PAWCS results [69]; row (4) shows SUBSENSE results [70]; row (5) shows CB results [71]; row (6) shows VUMETER results [72]; row (7) shows KDE results [73]; row (8) shows MULTICUE results [74]; row (9) shows SOBS results [75]; row (10) shows SC-SOBS results [76]; row (11) shows KNN results [77]; row (12) shows LBP-MRF results [78]; row (13) shows WMV results [79]; row (14) shows IDLM results [80]; row (15) shows MFCN results [81]; row (16) shows MSFgNet results [82]; row (17) shows ED-ResNet18 results [83].

Correct Classification (PCC) and Matthew's Correlation Coefficient (MCC). These evaluation metrics are measured based on four basic cardinalities: TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) and each of these cardinalities measure the amount of overlapping or missed area between the pixel level detection result and the binary ground-truth (i.e. object masks). A perfect detection is supposed to have higher values of F-Mes, PCC, and MCC (i.e., close to 1).

C. Experimental Results and Discussion

For objective evaluation, videos from each category of the atmospheric/weather conditions from the E-TUVD were considered. In our present study, E-TUVD video clips captured

TABLE V

PIXEL ORIENTED SIMPLE MATCHING COEFFICIENT BASED COMPARISON OF THE STATE-OF-THE-ART MOVING OBJECT DETECTION METHODS ON E-TUVD

State-of-the-Art Detection Methods		Pixel Oriented Similarity Measurement Metrics								
		Fog Condition			Haze Condition			Dust Condition		
		F-Mes (↑)	PCC (↑)	MCC (↑)	F-Mes (↑)	PCC (↑)	MCC (↑)	F-Mes (↑)	PCC (↑)	MCC (↑)
ViBe	[44]	0.4170	0.8492	0.4220	0.6493	0.9155	0.6744	0.3517	0.7687	0.3232
ISBM	[55]	**0.5700**	**0.8962**	**0.5973**	0.6873	0.9296	0.6932	0.4960	0.8378	0.4993
MTD	[56]	0.4377	0.8366	0.4458	0.4347	0.8257	0.4743	0.2818	0.6371	0.2413
PBAS	[57]	**0.5627**	**0.9001**	**0.5913**	0.6571	0.9158	0.6631	0.4743	0.8366	0.4724
GMM-V1	[58]	0.4747	0.8614	0.4884	0.6670	0.9244	0.6759	0.3543	0.7800	0.3546
BM	[59]	0.5446	0.8896	0.5306	**0.6952**	**0.9356**	**0.7026**	0.4999	0.8446	0.5001
XCS-LBP	[60]	***0.1468***	***0.6562***	***0.1561***	***0.2100***	***0.7027***	***0.2305***	0.3341	0.6895	0.2954
DECOLOR	[61]	0.3872	0.8313	0.4004	0.5720	0.9091	0.5775	0.3600	0.7752	0.3278
GMM-V2	[62]	0.5283	0.8822	0.5403	0.4383	0.8317	0.4881	0.2882	0.6893	0.2718
EB	[63]	0.4061	0.8255	0.4039	*0.6105*	*0.9048*	*0.6035*	0.4697	0.8469	0.4670
SC-RPCA	[64]	0.3772	0.8337	0.3629	0.4876	0.8442	0.5289	0.4652	0.8374	0.4508
IMTSL	[65]	0.4377	0.8420	0.4222	0.6129	0.9021	0.6044	0.4945	0.8403	0.4842
LOBSTER	[66]	0.4249	0.8492	0.4458	0.6196	0.9107	0.6092	0.5052	0.8423	0.5010
MBS	[67]	0.5387	0.8913	0.5302	0.5355	0.9013	0.5237	0.4923	0.8397	0.4687
SSS-RPCA	[68]	0.5402	0.8945	0.5351	0.6634	0.9176	0.6696	0.4967	0.8418	0.4822
PAWCS	[69]	0.4535	0.8589	0.4752	0.6368	0.9146	0.6253	**0.5492**	**0.8521**	**0.5436**
SUBSENSE	[70]	0.5564	0.8933	0.5450	0.6671	0.9204	0.6783	0.5195	0.8473	0.5161
CB	[71]	0.4943	0.8785	0.5298	0.5151	0.8732	0.5461	***0.2108***	***0.6137***	***0.2046***
VU-METER	[72]	0.4728	0.8584	0.4841	0.4578	0.8474	0.4889	0.3892	0.8173	0.3797
KDE	[73]	0.4399	0.8486	0.4434	0.6347	0.9133	0.6254	0.3803	0.8067	0.3558
MULTICUE	[74]	0.4344	0.8659	0.4601	0.5178	0.8539	0.5460	**0.5221**	**0.8510**	**0.5351**
SOBS	[75]	0.4838	0.8539	0.4694	0.6821	0.9251	0.6894	0.4426	0.8317	0.4402
SC-SOBS	[76]	0.4990	0.8795	0.5104	0.6803	0.9278	0.6861	0.4395	0.8361	0.4267
KNN	[77]	0.2074	0.7207	0.1715	0.6727	0.9197	0.6844	0.2399	0.6230	0.2203
LBP-MRF	[78]	0.3464	0.8197	0.3548	**0.6991**	**0.9317**	**0.6985**	0.5194	0.8372	0.5073
WMV	[79]	0.4826	0.8646	0.4937	0.6705	0.9223	0.6785	0.4572	0.8292	0.4140
IDLM	[80]	0.4906	0.8730	0.5298	0.5192	0.8941	0.5493	0.4294	0.8238	0.4046
MFCN	[81]	0.4885	0.8719	0.4981	0.5268	0.8973	0.5671	0.4128	0.8230	0.4138
MSFgNet	[82]	0.4795	0.8626	0.4726	0.5122	0.8859	0.5324	0.4389	0.8292	0.4125
ED-ResNet18	[83]	0.4663	0.8563	0.4927	0.5238	0.8834	0.5397	0.2453	0.6298	0.2415

State-of-the-Art Detection Methods		Pixel Oriented Similarity Measurement Metrics								
		Rain Condition			Poor Illumination Condition			Clear Day Condition		
		F-Mes (↑)	PCC (↑)	MCC (↑)	F-Mes (↑)	PCC (↑)	MCC (↑)	F-Mes (↑)	PCC (↑)	MCC (↑)
ViBe	[44]	0.4925	0.8827	0.4966	0.2854	0.7068	0.2856	0.6745	0.9450	0.6876
ISBM	[55]	**0.5347**	**0.8968**	**0.5689**	**0.5843**	**0.7597**	**0.5727**	0.6862	0.9568	0.6927
MTD	[56]	0.4871	0.8537	0.4951	0.1766	0.6268	0.2139	0.6046	0.9236	0.6156
PBAS	[57]	0.4796	0.8531	0.4869	0.4476	0.7469	0.4381	0.6895	0.9542	0.6876
GMM-V1	[58]	0.4514	0.8314	0.4652	0.4288	0.7383	0.4160	0.6764	0.9664	0.6964
BM	[59]	0.4562	0.8374	0.4662	0.3887	0.7268	0.3793	0.5437	0.9007	0.5884
XCS-LBP	[60]	***0.2167***	***0.6769***	***0.2294***	0.2261	0.6325	0.2205	***0.5233***	***0.9003***	***0.5634***
DECOLOR	[61]	**0.5347**	**0.8904**	**0.5689**	**0.4886**	**0.7494**	**0.4399**	**0.7735**	**0.9828**	**0.7758**
GMM-V2	[62]	0.5048	0.8755	0.4964	0.2861	0.7098	0.3011	0.6033	0.9200	0.6171
EB	[63]	0.4947	0.8819	0.5093	0.2604	0.6902	0.2313	0.6711	0.9598	0.6958
SC-RPCA	[64]	0.4798	0.8472	0.4865	0.2780	0.7023	0.2719	0.6560	0.9544	0.6875
IMTSL	[65]	0.4693	0.8595	0.4829	0.2339	0.6448	0.2245	0.7175	0.9811	0.7200
LOBSTER	[66]	0.4439	0.8271	0.4625	0.3382	0.7104	0.3171	0.6936	0.9789	0.7124
MBS	[67]	0.4298	0.8195	0.4547	0.2407	0.6895	0.2331	0.6905	0.9744	0.7012
SSS-RPCA	[68]	0.4365	0.8289	0.4602	0.2673	0.6934	0.2387	0.7069	0.9783	0.7168
PAWCS	[69]	0.2802	0.7683	0.2925	0.3561	0.7205	0.3742	0.7682	0.9792	0.7692
SUBSENSE	[70]	0.4246	0.8209	0.4337	0.3624	0.7110	0.3330	0.6511	0.9292	0.6796
CB	[71]	0.4653	0.8426	0.4820	0.1657	0.6042	0.1665	0.6171	0.9318	0.6425
VU-METER	[72]	0.4842	0.8552	0.4894	0.2404	0.6893	0.2352	0.6159	0.9248	0.6231
KDE	[73]	0.4880	0.8688	0.4962	***0.1330***	***0.5590***	***0.1282***	0.6962	0.9716	0.7059
MULTICUE	[74]	0.4276	0.8227	0.4543	0.3700	0.7107	0.3228	**0.7754**	**0.9834**	**0.7796**
SOBS	[75]	0.4905	0.8776	0.5343	0.4057	0.7292	0.3934	0.6226	0.9377	0.6582
SC-SOBS	[76]	0.4080	0.8011	0.4113	0.4063	0.7295	0.4149	0.6235	0.9329	0.6547
KNN	[77]	0.2294	0.6833	0.2573	0.1346	0.5893	0.1294	0.6014	0.9187	0.6212
LBP-MRF	[78]	0.4036	0.7980	0.4118	0.1388	0.5980	0.1386	0.5792	0.9166	0.5892
WMV	[79]	0.4173	0.8215	0.4122	0.3285	0.7086	0.3150	0.6286	0.9322	0.6582
IDLM	[80]	0.2479	0.7009	0.2896	0.1733	0.6232	0.2118	0.5925	0.9174	0.6099
MFCN	[81]	0.3872	0.7443	0.4052	0.1678	0.6201	0.2046	0.6199	0.9277	0.6475
MSFgNet	[82]	0.2758	0.7854	0.3444	0.1714	0.6225	0.2108	0.6093	0.9213	0.6279
ED-ResNet18	[83]	0.2367	0.6973	0.2748	0.1583	0.5946	0.1659	0.5938	0.9182	0.6121

Bold Face and Underlined: Outer Performed Method; Bold Face: Second Most Outer Performed Method; Bold Face and Italic: Lowest Performed Method; ↑- Higher Values Indicate Perfect Segmentation; ↓- Lower Values Indicate Perfect Segmentation

under the static background conditions were used. For equitable decision making regarding the effective and efficient state-of-the-art moving object detection methods, all the aforementioned metrics were measured for each detection method. Each of the methods was tested on a CPU platform workstation with Intel®Xeon®Processor E5-1620 v3 @ 3.50 GHz and 64 GB installed memory (RAM). The average performance of each detection methods over each video category of the E-TUVD has been tabulated in TABLE V. The two best performing methods in each of the representative

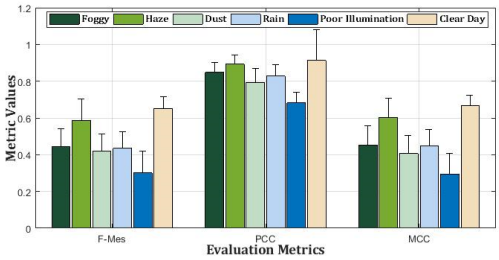

Fig. 5. Error bar plot of assessment metrics for all moving object detection methods over each category of weather/atmospheric conditions of E-TUVD.

atmospheric/weather challenges are presented in boldface with underline and boldface. Conversely, the method with the worst performance is presented in italics and boldface. Careful inspection of TABLE V provides the following observations:

1) The top two outer-performed methods in each of the video categories are not necessarily the best methods for all the considered video categories. Higher values were achieved for all evaluation metrics using the following methods: fog condition in ISBM [55] and PBAS [57]; haze condition in BM [59] and LBP-MRF [78]; dust condition in PAWCS [69] and MULTICUE [74]; rain and poor illumination conditions in DECOLOR [61] and ISBM [55] and clear day conditions in DECOLOR [61] and MULTICUE [74]. These can be considered the two outer-performed methods in each of the respective video category. Conversely, XCS-LBP [60] was observed to be the worst performing method in almost all video categories, except dust and poor illumination conditions. For dust and poor illumination conditions, CB [71] and KDE [73] showed lower F-Mes, PCC, and MCC values and can be considered the worst performing methods, respectively.
2) Fig. 5 shows the error bar plot of performance evaluation metrics for each video category over all the used methods. In this plot, each bar represents the variability of all detection methods over each video category to indicate the error or uncertainty in a reported assessment metric. As can be seen from Fig. 5, it should come no surprise that for all the assessment metrics, videos captured in clear day conditions exhibit high value for all three assessment metrics (i.e., F-Measure (0.6502), PCC (0.9440), and MCC (0.6679)) for all the methods. In contrast, almost all methods faced difficulties with videos in poor illumination conditions because they mostly suffered from camouflage problems, resulting in lower values of F-Mes (0.2881), PCC (0.6779), and MCC (0.2843).
3) Furthermore, video clips of E-TUVD under fog, haze and dust conditions were also challenging categories and provided low metric values as compared to the clear day situation. This is due to the effect of reduced visibility on the ability to automatically detect moving objects from long-distance video sequences degraded by the atmospheric path. Moreover, detection algorithms based on deep learning approaches (i.e., IDLM [80], MFCN [81], MSFgNet [82], and ED-ResNet18 [83]) used in our study did not show remarkable results in any of the video categories of the E-TUVD except on clear day conditions.

VI. Visibility Enhancement of Degraded Atmospheric/Weather Image Sequences Using State-of-the-Art Methods

Degraded atmospheric/weather conditions drastically alter images of natural outdoor scenarios. Under such conditions, the performance of any moving object detection algorithm deteriorates (as can be observed from TABLE V). Thus, it is necessary to enhance and remove the visual effects of these atmospheric/weather conditions to improve the performance of detection algorithms. In this section, we compare the state-of-the-art visibility enhancement methods for restoration of atmospheric/weather degraded image sequences (under fog, haze and dust conditions) of the E-TUVD.

A. State-of-the-Art Methods Tested

Over the years, many algorithms have been proposed for visibility enhancement and restoration of real-world images for various vision based applications. The most representative visibility enhancement techniques can be based on the input information:

- **Multiple Image Approaches:** Atmospheric Scattering Model ([42], [43], [84]), Depth Estimation ([85]–[87]), and Polarizing Filtering ([88]–[90]).
- **Single Image Approaches:** Fusion strategy ([91]–[94]), Dark Channel Prior ([95]–[98]), Retinex Theory ([99]–[103]), Bayesian Strategy ([104], [105]), Filtering ([106]–[109]); Histogram Equalization ([110]–[114]), Learning Strategy ([115], [116]), and Deep Learning ([117]–[121]).

These approaches are used as a basic framework to develop algorithms for visibility restoration of outdoor scenes. The multiple image based visibility enhancement techniques requires additional information such as elevation, camera position, and the approximate distance between the view point and the sky area of the captured images. Unfortunately, these techniques often suffer from various issues that reduce the practical applicability of these techniques. In contrast, nowadays, most studies have focused on single image approaches because these methods usually approximate the thickness of atmospheric conditions from a single input image. From the literature, a total of 22 single image based visibility enhancement methods (i.e., 20 classical methods and 2 deep learning methods) were implemented and tested on atmospheric/weather degraded image sequences of the E-TUVD. For each method, the parameters were adjusted based on a subjective measurement of the overall results, and one set of parameters was used for all the videos used in our analysis. A brief overview of these visibility enhancement techniques, along with the associated parameters tuning is provided in TABLE VI.

Benchmarking these methods, is advantageous because these methods can be quantitatively compared for the restoration of atmospheric/weather degraded outdoor images

TABLE VI

BRIEF SUMMARY OF THE VISIBILITY ENHANCEMENT TECHNIQUES USED IN THE E-TUVD DATASET WITH THE ASSOCIATED PARAMETERS

Method ID	Method Name		Approach/ Technique	IC	Parameters Setting
MSF	Multi Scale Fusion	[93]	SI/ FBS	M	n- 6; γ- 2.5; σ- 0.3; S_{max}- 1; W_{hc}- π/2.75; L- 5; d- 16
DCP	Dark Channel Prior	[95]	SI/ DCP	M	S_F - 0.5; ω- 0.95; λ- 10^{-4}; W_s- 3; €- 10^{-8}; t_0- 0.1; P_s- 15, P- 7
FVR	Fast Visibility Restoration	[109]	SI/ FI	M	V_M- 300; S_v- 51; p- 0.95; W_B- 0.5; S_{max}- 1; G_F- 1.3; V_H- 0.1
STD	Structure Texture Decomposition	[122]	SI/ DCP	M	T- 0.7; λ- 10^5; β_L- 0.5
BD	Bayesian Defogging	[105]	SI/ BS	M	C- 2
BCCR	Boundary Constraint and Contextual Regularization	[98]	SI/ DCP	M	W_s- 10; C0- 10; C1- 40; λ- 2; C- 0.0001; σ- 0.2; α_s- 0.5; δ- 0.6
SE	Stochastic Enhancement	[123]	SI/ FI	M	C_p- 0.1
CAP	Color Attenuation Prior	[124]	SI / DCP	P	r- 15; β- 1.0; θ_0- 0.12; θ_1 - 0.95; θ2- -0.78; σ- 0.04; F_s- 60; €- 10^{-6}; t_0- 0.05; t_1- 1
MSCNN	Multi-Scale Convolutional Neural Networks	[117]	SI/ DL	M	W_B- 1; T- 1; G_F-1.7; W_S- 15
NLR	Non Local Prior	[125]	SI/ PLOD	M	λ_L- 0.1; G_F- 1.3; A_{min}/A_{max} - 1
GIF	Guided Image Filtering	[107]	SI/ FI	M	r- 60; β- 1.0; F_s- 60; €- 10^{-1}
VAS	Visual Artifact Suppression	[126]	SI/ PLOD	M	t_0- 0.2; β_S- 6; γ_M- 0:77; α_0- 0.5; α_1- 0.05; λ- 0.01; η- 0.1
OTM	Optimal Transmission Map	[127]	SI/ PLOD	M	β- 0.8; G_F- 0.5; P_W- 1; €- 10^{-5}, M_C- 10^8
IPR	Internal Patch Recurrence	[128]	SI/ PLOD	M	NN- 9; S_F- 0.95; P_R- 5e-3; λ- 0.5; P_S- 7
MLP	Multilayer Perceptron	[115]	SI/ LS	M	W_s- 16; δ_S- 8; ω_L- 0.7; t_0- 0.1
CO	Convex Optimization	[129]	SI/ PLOD	M	G_F- 1.44; λ_1- 0.02; λ_2- 0.002; λ_3- 0.04
OCE	Optimized Contrast Enhancement	[130]	SI/ HE	M	B_S- 100; P_S- 8; λ_L- 5; σ_V- 10; t_0- 0.1; G_F- 0.8; €- 10^{-6}
CBF	Color Balance and Fusion	[92]	SI/ FBS	M	SL_1- 0.001; SL_2- 0.005; μ- 0.5; σ- 0.25; λ- 0.1; L- 5
RB	Retinex Based	[100]	SI/ RBS	M	μ_I- 2.3; α_P- 100; β_P- 0.1; γ_P- 1; λ- 10
GPR	Two-layer Gaussian Process Regression	[131]	SI/	M	SP_C- 200; SP_F- 1000; EV- 50; S_R- 0.2; S- -0.005; t_0- 0.1; r- 35; €- 10^{-3}; S_C- 0.2; δ- 0.2
DehazeNet	Dehazing Network	[119]	SI/ DL	P	r- 50; €- 10^{-3}
CLAHE	Contrast-limited adaptive histogram equalization	[110]	SI/ HE	M	W_s- 8; C_{EL}- 0.008; α- 0.6

IC- Implemented Code; SI- Single Image Approach; FBS- Fusion Based Strategy; DCP- Dark Channel Prior Based Strategy; FI- Filtering Based Strategy; BS- Bayesian Based Strategy; DL- Deep Learning Based Strategy; LS- Learning Based Strategy; HE- Histogram Equalization Based Strategy; RBS- Retinex Based Strategy; n- norm; γ- Luminance factor; μ- Average Constant; σ- Standard Deviation; S_{max}- Higher Saturated Pixel Constant; W_{hc}- Higher Frequency Cut-off; L- Pyramid Level; d- Upsampling Factor; S_F- Scaling Factor; €/λ/ $\lambda_1/\lambda_2/\lambda_3/\alpha0/\alpha_1$- Regularization Parameter; W_s- Window Size; t_0- Lower Bound of Transmission; P_s- Patch Size; P- Pad Size; V_M- Minimum Visibility Observable Distance; S_v- Segment Length; p- Percentage of Restoration; W_B- White Balance; G_F- Gamma Correction Factor; V_H- Value for the Height of the Horizon Line; T- Threshold Parameter; λ_L/β_L- Smoothness Level Controller; ω/ C- Constant/ Multiplicative Constant; C0/C1- Boundary Constants ;C_{EL}- Contrast Enhancement Limit; α- Distribution Parameter; α_s- Sensitivity Controller; δ- Exponential Constant; C_p- Small Positive Constant; r- Local window Radius; β- Scattering Coefficient; $\theta_0/\theta_1/\theta2$- Linear Coefficients; F_s- Filter Size; t_0/t_1- Minimum and Maximum Bound of Transmission; W_B- White Balance; T- Threshold; A_{min}/ A_{max}- Minimal and Maximal Value of Airlight; D_R- Dehazing Ratio; β_S- Sharpness Adjustable Parameter; γ_M- Magnitude Adjustable Parameter; η- Weighting Parameter; M_C- Maximum Pixels for Coarse Scale; δ_S- Window Sliding Value; ω_L- Desired Restoration Level; B_S- Block Size; λ_L- Weighting Parameter; σ_V- Controls the Variance of Probability; SL_1/ SL_2- Saturation Level; μ_I- Parameter to control Image Dynamics; $\alpha_P/\beta_P/\gamma_P$- Free Positive Parameters; SP_C/SP_F- Number of Super Pixels Coarse/ Fine; EV- Number of Eigen Vectors; S_R- Sampling Rate; S- Stability/ Proximity; S_C- Sky Compensation; SF- Scaling Factor; PR- Global Pruning Rate

(i.e., owing to fog, haze and dust). Compared with the corresponding original frames, the results of these visibility enhancement methods on E-TUVD are displayed in Fig. 6.

B. Qualitative Assessment Metrics

Measuring the perceptual quality of images by objective quality assessment metrics that agree with the human observer is a fundamental need of image processing. In case of real-world applications, reference images (i.e., ground-truth images) are not available, and the quality evaluation is solely based on the test images. Hence we performed quantitative evaluation using blind measures (i.e., no-reference quality assessment metrics were used) [132]. The metrics considered in our study are: rate of new visible edges (e), mean ratio (r), percentage of saturated pixels (σ), and visibility measurement (VM). The first two metrics (e and r) use the enhanced degree of image edges to approximate the enhanced degree of image visibility. The third metric (σ) indicates color restoration performance of enhancement algorithms by approximating the rate of saturated pixels after enhancement. Finally, the last metric (VM) quantifies the degree of the visibility of the image after enhancement based on visible edge segmentation. A well enhanced image is supposed to have high values of e, r, and VM and a low value of σ.

C. Experimental Results and Discussion

For qualitative comparison, similar videos used for testing the detection algorithms in fog, haze and dust conditions were selected from the E-TUVD dataset. To compute the simulation results, visibility enhancement algorithms were implemented on a CPU platform of a 64-bit workstation as mentioned in Section V. The average values of the aforesaid assessment metrics for restoring the visibility of atmospheric/weather degraded image sequences are reported in TABLE VII. Similar to TABLE V, the two best performance methods in each of the representative atmospheric/weather challenges are presented in boldface with underline and boldface. Conversely, the worst performance methods are presented in italics and boldface. TABLE VII and Fig. 6, show that the quality assessment indexes may not be absolutely consistent with the subjective assessment, but they can be used as references for comparing and benchmarking different enhancement algorithms. The following observations can be made:

1) For fog and haze conditions, the SE method [123] achieved superior values for all the visibility assessment criteria and hence can be considered the best performance method among all methods in preserving both the colors and edges and effectively enhancing the visibility. Although the SE method [123] provided effective results for these two aforementioned conditions, it did not effectively enhancing visibility of image sequences degraded because of dust conditions and exhibited color distortion.
2) Moreover, no single enhancement algorithm showed the best performance under the considered atmospheric/ weather conditions (i.e., degraded due to fog, haze, and dust). For dust conditions, the CBF method [92] and NLR method [125] were the two outer-performed methods. Additionally, the NLR method [125] also effectively restored the visibility of scenes under fog conditions and can be considered the second most outer-performed method after the SE method [123].
3) The DCP method proposed by He et.al. [95] is one of the most popular single image based enhancement

Fig. 6. Visual comparison of the visibility enhancement techniques on E-TUVD: **First row** shows Input Frames; **Second row**: Col (1,3,5) shows MSF [93] results; Col (2,4,6) shows DCP [95] results; **Third row**: Col (1,3,5) shows FVR [109] results; Col (2,4,6) shows STD [122] results; **Fourth row**: Col (1,3,5) shows BD [105] results; Col (2,4,6) shows BCCR [98] results; **Fifth row**: Col (1,3,5) shows SE [123] results; Col (2,4,6) shows CAP [124] results; **Sixth row**: Col (1,3,5) shows MSCNN [117] results; Col (2,4,6) shows NLR [125] results; **Seventh row**: Col (1,3,5) shows GIF [107] results; Col (2,4,6) shows VAS [126] results; **Eight row**: Col (1,3,5) shows OTM [127] results; Col (2,4,6) shows IPR [128] results; **Ninth row**: Col (1,3,5) shows MLP [115] results; Col (2,4,6) shows CO [129] results; **Tenth row**: Col (1,3,5) shows OCE [130] results; Col (2,4,6) shows CBF [92] results; **Eleventh row**: Col (1,3,5) shows RB [100] results; Col (2,4,6) shows GPR [131] results; **Twelfth row**: Col (1,3,5) shows DehazeNet [119] results; Col (2,4,6) shows CLAHE [110] results.

algorithms, and in recent years, various algorithms have been proposed depending on this concept. To improve the edge preservation and efficiency, the improved DCP algorithm as proposed by Meng et.al. [98] i.e. the BCCR method, exhibited larger values of all visibility criteria than the algorithm proposed by He et.al. [95] and could comparatively improve efficiency.

4) Compared with all enhancement algorithms, the STD [122], IPR [128], and CLAHE [110] methods exhibited smaller values in terms of qualitative assessment metrics for all the three degraded atmospheric/ weather conditions (i.e., fog, haze and dust), and these results were consistent with the subjective assessment, as shown in Fig. 6.

TABLE VII

NO-REFERENCE IMAGE BASED QUALITATIVE EVALUATION OF THE STATE-OF-THE-ART VISIBILITY ENHANCEMENT METHODS ON E-TUVD

State-of-the-Art Enhancement Methods		No-Reference Quality Assessment											
		Fog Condition Dataset				Haze Condition Dataset				Dust Condition Dataset			
		e (↑)	r (↑)	σ (↓)	VM (↑)	e (↑)	r (↑)	σ (↓)	VM (↑)	e (↑)	r (↑)	σ (↓)	VM (↑)
MSF	[93]	0.4573	1.1822	0.0688	7.9677	**0.9507**	**2.0454**	**0.0046**	**17.8229**	0.6539	1.8131	0.0081	11.2883
DCP	[95]	0.4669	1.2027	0.0666	7.6727	0.6003	1.7388	0.0132	14.0345	0.4880	1.3241	0.0492	10.4237
FVR	[109]	0.3369	1.1629	0.0761	7.5450	0.4348	1.3383	0.0266	15.3442	0.2787	0.8824	0.0983	9.5024
STD	[122]	***0.2177***	***1.0356***	***0.5521***	***5.6153***	0.3677	1.0842	0.0400	14.8770	0.5398	1.5873	0.0309	11.5771
BD	[105]	0.4325	1.1763	0.0773	7.7496	0.4714	1.3801	0.0224	14.0696	0.1604	0.6531	0.0827	7.6309
BCCR	[98]	0.8788	1.6445	0.0319	9.8475	0.5253	1.5769	0.0167	14.0613	0.4637	1.0701	0.0577	10.2269
SE	[123]	**0.9277**	**2.4648**	**0.0055**	**11.152**	**0.9633**	**2.4526**	**0.0013**	**19.2568**	0.3107	0.9496	0.0800	9.4016
CAP	[124]	0.3096	1.1512	0.1056	7.1514	0.5230	1.5269	0.0172	14.2355	0.3373	0.9470	0.0750	9.6345
MSCNN	[117]	0.8649	1.1822	0.0268	9.0604	0.6532	1.9007	0.0100	15.9600	0.6602	1.4088	0.0275	10.5201
NLR	[125]	**0.9197**	**1.9977**	**0.0117**	**9.9659**	0.9268	1.9779	0.0057	17.3002	**0.7975**	**1.9077**	**0.0004**	**12.3979**
GIF	[107]	0.7253	1.4639	0.0414	8.9985	0.6449	1.8876	0.0090	15.3614	0.4185	1.0421	0.0602	10.6763
VAS	[126]	0.5235	1.2466	0.0598	8.7318	0.8798	1.9220	0.0079	16.5669	0.4122	1.0008	0.0632	9.6756
OTM	[127]	0.5460	1.2684	0.0623	8.0198	0.7889	1.9087	0.0081	16.2680	0.2258	0.8557	0.0884	8.2912
IPR	[128]	0.2528	1.0896	0.4053	6.7403	0.3395	1.0890	0.0642	12.5370	***0.1213***	***0.0894***	***0.1555***	***6.2763***
MLP	[115]	0.7787	1.4872	0.0341	9.0475	0.6521	1.8352	0.0076	16.4482	0.5510	1.7027	0.0200	11.9143
CO	[129]	0.4942	1.2481	0.0501	7.7191	0.5042	1.4666	0.0211	14.1601	0.4341	1.1622	0.0495	10.6526
OCE	[130]	0.7233	1.4037	0.0416	8.5341	0.6125	1.8184	0.0105	13.1381	0.2279	0.8824	0.0893	8.5973
CBF	[92]	0.7542	1.4552	0.0351	8.1263	0.3827	1.1417	0.0387	14.6075	**0.8782**	**1.9201**	**0.0076**	**14.4955**
RB	[100]	0.6586	1.3364	0.0453	8.6802	0.3958	1.1623	0.0313	14.3570	0.5049	1.3747	0.0329	9.3227
GPR	[131]	0.5313	1.2548	0.0560	8.9256	0.3923	1.3322	0.0368	14.6611	0.3552	0.9755	0.0740	9.2677
DehazeNet	[119]	0.5985	1.2931	0.0458	8.2933	0.6171	1.7495	0.0105	15.2665	0.3681	0.9687	0.0637	10.1492
CLAHE	[110]	0.2881	1.1414	0.3728	6.2371	***0.3024***	***1.0595***	***0.1100***	***11.1988***	0.1564	0.5018	0.1022	7.1757

Bold Face and Underlined: Outer Performed Method; Bold Face: Second Most Outer Performed Method; Bold and Italic Face: Lowest Performed Method; ↑- Higher Values Indicate Better Enhancement; ↓- Lower Values Indicate Better Enhancement

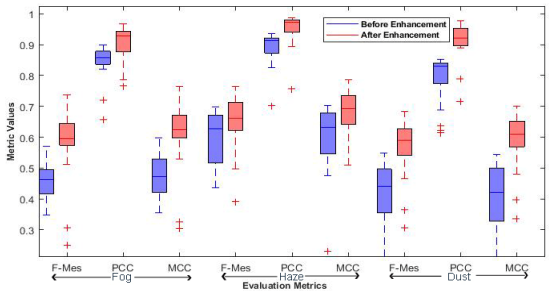

Fig. 7. Box-whisker plot of the assessment metrics for moving object detection methods in pre-degradation (i.e. before visibility enhancement represented by blue boxes) and post restoration (i.e. after visibility enhancement represented by red boxes) of degraded atmospheric/weather conditions of E-TUVD.

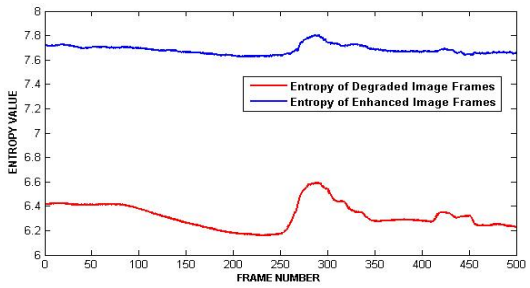

Fig. 8. Characterization of the visibility enhanced frames from the E-TUVD dataset using Entropy (For clarity the Y-axis has been shown for range [6-8]).

VII. INFLUENCE OF VISIBILITY ENHANCEMENT FOR ACCURATE DETECTION OF MOVING OBJECTS IN DEGRADED ATMOSPHERIC/ WEATHER CONDITIONS

In this section, we demonstrate the practical utility of visibility enhancement for accurate detection of moving objects in atmospheric/weather degraded outdoor scenes. To validate this point, considering the most outer-performed visibility enhancement method (i.e., the SE method [123] for fog and haze conditions and the CBF method [92] for dust conditions) in TABLE VII, Fig. 7 displays the box-whisker plot of pre-degradation (i.e., represented by blue boxes) and post restoration (i.e., represented by red boxes) performance results of moving object detection methods (as described in TABLE V) of outdoor scenes. In this non-parametric plot, each box is enclosed by first and third quartiles to represent groups of numerical data and is divided into two parts by the median value represented using red lines on the Y-axis. The outliers (represented by red asterisks) in the plot show the extreme variability of the assessment metrics in the dataset. Because of the limitation of space, we cannot display the experimental results of moving object detection algorithms after visibility restoration in a tabulated form. Fig. 7 shows that successful image restoration as a pre-processing step may significantly improve the performance of moving object detection algorithms (an average accuracy of 5.82 and 6.08 percentage-point improvement under fog and haze conditions and 11.37 percentage-point improvement under dust conditions). To characterize the texture information in pre-degraded and post-restored frames of the E-TUVD, we used entropy to measure the contents, as shown in Fig. 8. The higher entropy values of restored frames over those of the degraded frames (shown in Fig. 8) indicate an image with adequate details, thereby denoting good quality. Therefore, the distinguishable information of enhanced frames can reveal hidden salient objects as compared to the atmospheric/weather degraded frames. For the same reason, the detection results obtained in our case after restoring the atmospheric/weather degraded image sequences are higher than those obtained before restoration. Some significant improvements can also be observed in Fig. 9, which shows that moving object detection after visibility restoration suppresses

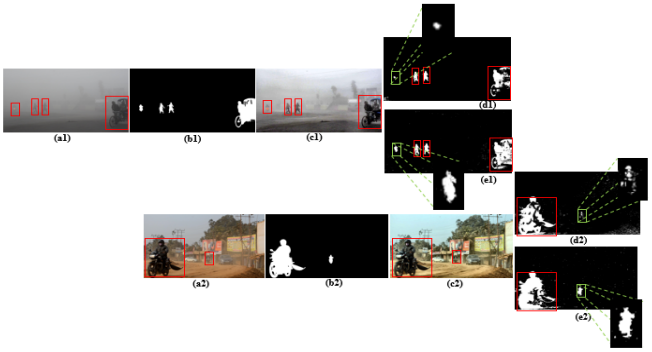

Fig. 9. Results of applying the moving object detection algorithm to the post restored and pre-degraded video sequences: (a1),(a2): input frames in fog and dust condition; (b1),(b2): Corresponding Ground-truth frames; (c1),(c2) visibility enhanced frames (SE result [123] for fog condition and CBF result [92] for Dust condition); (d1), (d2): Moving object detection before enhancement (ISBM result [55] for fog condition and PAWCS result [69] for dust condition); (e1),(e2): Moving object detection after enhancement (ISBM result [55] for fog condition and PAWCS result [69] for dust condition).

all false detections, while keeping the "real" moving objects. In addition, the shapes of the extracted moving objects appeared more clearly in the restored frames than in the atmospheric/weather degraded image sequences. Fig. 9(e1) clearly shows that the moving objects detected are two walking human beings, one scooter rider, and one rickshaw puller after visibility enhancement; in contrast, in Fig. 9(d1), the scooter rider (far distance object) is not well detected because of fog. Similarly, in Fig. 9(e2), the moving objects detected are two bike riders, whereas in the corresponding frame which is degraded because of dust (Fig. 9(d2)), it is difficult make this conclusion. This emphasizes the benefit of visibility restoration as a basic tool for high level target acquisition (such as recognition) for both human visual systems and computerized applications.

VIII. Conclusion

We present a ground-truth annotated benchmark video dataset named as Extended Tripura University Video Dataset (E-TUVD) for moving object detection in atmospheric/weather degraded outdoor scenes. The dataset aims to provide the research community with a benchmark facility for moving object detection under dynamic variations of adverse atmospheric/weather conditions in an outdoor environment. We then investigate the potentiality of several popular state-of-the-art moving object detection and visibility enhancement techniques on the E-TUVD dataset. Furthermore, the effect of visibility restoration on the ability to automatically acquire moving objects was also examined here for video sequences degraded by extreme atmospheric/weather conditions. Although there are still many limitations that need to be addressed further, the purpose of this study has been successfully fulfilled. The dataset will be extended to include other atmospheric/weather conditions and will be regularly revised based on feedback from the research community. In the future, our aim is to design an unsupervised representation learning model based on deep convolutional neural networks for accurate detection of moving objects in degraded atmospheric/weather conditions to improve the performance of each evaluation metric and to get the detection results as close as possible to the ground-truth images.

Acknowledgment

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

References

[1] X. Ji, Z. Wei, and Y. Feng, "Effective vehicle detection technique for traffic surveillance systems," *J. Vis. Commun. Image Represent.*, vol. 17, no. 3, pp. 647–658, Jun. 2006.
[2] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabaly, and C. Quek, "Challenges in video based object detection in maritime scenario using computer vision," in *Proc. 19th Int. Conf. Connected Vehicles*, Jul. 2017, pp. 1–6.
[3] B. Garcia-Garcia, T. Bouwmans, and A. J. Rosales Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Comput. Sci. Rev.*, vol. 35, Feb. 2020, Art. no. 100204.
[4] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "A novel video dataset for change detection benchmarking," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4663–4679, Nov. 2014.
[5] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequièvre, "A benchmark dataset for outdoor foreground/background extraction," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 291–300.
[6] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," in *Proc. 12th IEEE Int. Workshop Perform. Eval. Tracking Surveill.*, Dec. 2009, pp. 1–6.
[7] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.
[8] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "ETISEO, performance evaluation for video surveillance systems," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 476–481.
[9] *VSSN 2006 Test Images Sequences*. Accessed: Feb. 23, 2019. [Online]. Available: http://mmc36.informatik.uni-augsburg.de/VSSN06_OSAC/
[10] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. V. Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 724–732.
[11] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 255–261.
[12] W. Wang, J. Shen, and L. Shao, "Consistent video saliency using local gradient flow optimization and global refinement," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4185–4196, Nov. 2015.
[13] *CAVIAR Test Case Scenarios*. Accessed: Apr. 4, 2019. [Online]. Available: http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/
[14] D. Tsai, M. Flagg, A. Nakazawa, and J. M. Rehg, "Motion coherent tracking using multi-label MRF optimization," *Int. J. Comput. Vis.*, vol. 100, no. 2, pp. 190–202, Nov. 2012.
[15] F. Li, T. Kim, A. Humayun, D. Tsai, and J. M. Rehg, "Video segmentation by tracking many figure-ground segments," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2192–2199.
[16] P. Ochs, J. Malik, and T. Brox, "Segmentation of moving objects by long term video analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1187–1200, Jun. 2014.
[17] J. Li, C. Xia, and X. Chen, "A benchmark dataset and saliency-guided stacked autoencoders for video-based salient object detection," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 349–364, Jan. 2018.
[18] I. Kavasidis, S. Palazzo, R. D. Salvo, D. Giordano, and C. Spampinato, "An innovative Web-based collaborative platform for video annotation," *Multimedia Tools Appl.*, vol. 70, no. 1, pp. 413–432, May 2014.
[19] R. Vezzani and R. Cucchiara, "Video surveillance online repository (ViSOR): An integrated framework," *Multimedia Tools Appl.*, vol. 50, no. 2, pp. 359–380, Nov. 2010.
[20] S. Blunsden and R. B. Fisher, "The BEHAVE video dataset: Ground truthed video for multi-person behavior classification," *Ann. BMVA*, vol. 4, vol. 4, pp. 1–12, 2010.

[21] D. Bloisi, L. Iocchi, M. Fiorini, and G. Graziano, "Camera based target recognition for maritime awareness," in *Proc. 15th Int. Conf. Inf. Fusion*, Jul. 2012, pp. 1982–1987.

[22] H. Office Scientific Development Branch, "Imagery library for intelligent detection systems (i-LIDS)," in *Proc. IET Conf. Crime Secur.*, 2006, pp. 445–448.

[23] M. Camplani, L. Maddalena, G. M. Alcover, A. Petrosino, and L. Salgado, "A benchmarking framework for background subtraction in RGBD videos," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2017, pp. 219–229.

[24] T. Bouwmans, L. Maddalena, and A. Petrosino, "Scene background initialization: A taxonomy," *Pattern Recognit. Lett.*, vol. 96, pp. 3–11, Sep. 2017.

[25] C. Cuevas, E. M. Yáñez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA," *Comput. Vis. Image Understand.*, vol. 152, pp. 103–117, Nov. 2016.

[26] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[27] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 918–923, Jul. 2003.

[28] *Laboratory for Image and Media Understanding Dataset (LIMU)*. Accessed: Feb. 21, 2020. [Online]. Available: http://limu.ait.kyushu-u.ac.jp/dataset/en/

[29] C. Benedek and T. Sziranyi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 608–621, Apr. 2008.

[30] R. Collins, X. Zhou, and S. K. Teh, "An open source tracking testbed and evaluation Web site," in *Proc. IEEE PETS Workshops*, vol. 35, Jan. 2005, pp. 1–8.

[31] S. Li, D. Florencio, Y. Zhao, C. Cook, and W. Li, "Foreground detection in camouflaged scenes," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4247–4251.

[32] C. Li, X. Wang, L. Zhang, J. Tang, H. Wu, and L. Lin, "Weighted low-rank decomposition for robust grayscale-thermal foreground detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 725–738, Apr. 2017.

[33] 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 *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2936–2940.

[34] G. Yao, T. Lei, J. Zhong, P. Jiang, and W. Jia, "Comparative evaluation of background subtraction algorithms in remote scene videos captured by MWIR sensors," *Sensors*, vol. 17, no. 9, p. 1945, 2017.

[35] G. Moyà-Alcover, A. Elgammal, A. Jaume-i-Capó, and J. Varona, "Modeling depth for nonparametric foreground segmentation using RGBD devices," *Pattern Recognit. Lett.*, vol. 96, pp. 76–85, Sep. 2017.

[36] P.-M. Jodoin, L. Maddalena, A. Petrosino, and Y. Wang, "Extensive benchmark and survey of modeling methods for scene background initialization," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5244–5256, Nov. 2017.

[37] S. Abdelhedi, A. Wali, and A. M. Alimi, "Toward a kindergarten video surveillance system (KVSS) using background subtraction based Type-2 FGMM model," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit. (SoCPaR)*, Aug. 2014, pp. 440–446.

[38] *Underwater Change Detection Dataset*. Accessed: Feb. 21, 2020. [Online]. Available: http://underwaterchangedetection.eu/index.html

[39] R. Kalsotra and S. Arora, "A comprehensive survey of video datasets for background subtraction," *IEEE Access*, vol. 7, pp. 59143–59171, 2019.

[40] S. D. Roy, M. K. Bhowmik, and J. Oakley, "A ground truth annotated video dataset for moving object detection in degraded atmospheric outdoor scenes," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1318–1322.

[41] *Extended Tripura University Video Dataset (E-TUVD)*. Accessed: May 15, 2019. [Online]. Available: http://mkbhowmik.in/eTuvd.aspx

[42] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[43] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

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

[45] *LabelImg*. Accessed: Feb. 11, 2020. [Online]. Available: https://github.com/tzutalin/labelImg

[46] C. Cuevas, E. Yáñez, and N. García, "Tool for semiautomatic labeling of moving objects in video sequences: TSLAB," *Sensors*, vol. 15, no. 7, pp. 15159–15178, 2015.

[47] TSLAB: Tool for Semiautomatic LABeling [Online]. [Online]. Available: https://www.gti.ssr.upm.es/data/TSLAB.html

[48] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.

[49] C. Cuevas, R. Martínez, and N. García, "Detection of stationary foreground objects: A survey," *Comput. Vis. Image Understand.*, vol. 152, pp. 41–57, Nov. 2016.

[50] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Comput. Sci. Rev.*, vol. 28, pp. 157–177, May 2018.

[51] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vols. 11–12, pp. 31–66, May 2014.

[52] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[53] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung, "Deep neural network concepts for background subtraction: A systematic review and comparative evaluation," *Neural Netw.*, vol. 117, pp. 8–66, Sep. 2019.

[54] A. Sobral and T. Bouwmans, "BGS library: A library framework for algorithm's evaluation in foreground/background segmentation," in *Background Modeling and Foreground Detection for Video Surveillance*. Boca Raton, FL, USA: CRC Press, 2014, pp. 559–574.

[55] F. C. Cheng, S. C. Huang, and S. J. Ruan, "Illumination-sensitive background modeling approach for accurate moving object detection," *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 794–801, Dec. 2011.

[56] W. H. Lee, "Foreground objects detection using multiple difference images," *Opt. Eng.*, vol. 49, no. 4, Apr. 2010, Art. no. 047201.

[57] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38–43.

[58] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, pp. 246–252.

[59] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.

[60] C. Silva, T. Bouwmans, and C. Frélicot, "An eXtended center-symmetric local binary pattern for background modeling and subtraction in videos," in *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.*, 2015, pp. 1–8.

[61] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.

[62] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, Aug. 2004, pp. 28–31.

[63] Y. Tian, Y. Wang, Z. Hu, and T. Huang, "Selective eigenbackground for background modeling and subtraction in crowded scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1849–1864, Nov. 2013.

[64] Y. Li, G. Liu, and S. Chen, "Detection of moving object in dynamic background using Gaussian max-pooling and segmentation constrained RPCA," 2017, *arXiv:1709.00657*. [Online]. Available: http://arxiv.org/abs/1709.00657

[65] A. Sobral, C. G. Baker, T. Bouwman, and E. H. Zahzah, "Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2014, pp. 94–103.

[66] P. L. St-Charles and G.-A. Bilodeau, "Improving background subtraction using local binary similarity patterns," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2014, pp. 509–515.

[67] H. Sajid and S. C. S. Cheung, "Universal multimode background subtraction," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3249–3260, Jul. 2017.

[68] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 1007–1022, Feb. 2019.

[69] P. L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Universal background subtraction using word consensus models," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4768–4781, Oct. 2016.

[70] P. L. S. Charles, G. A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[71] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 5, Oct. 2004, pp. 3061–3064.

[72] Y. Goyat, T. Chateau, L. Malaterre, and L. Trassoudaine, "Vehicle trajectories evaluation by static video sensors," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 864–869.

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

[74] 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. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 527–540, Jun. 2012.

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

[76] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 21–26.

[77] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.

[78] C. Kertész, "Texture-based foreground detection," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 4, no. 4, pp. 51–62, 2011.

[79] D. Tian, H. Mansour, and A. Vetro, "Depth-weighted group-wise principal component analysis for video foreground/background separation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3230–3234.

[80] Y. Wang, Z. Luo, and P. M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognit. Lett.*, vol. 96, pp. 66–75, Sep. 2017.

[81] D. Zeng and M. Zhu, "Background subtraction using multiscale fully convolutional network," *IEEE Access*, vol. 6, pp. 16010–16021, 2018.

[82] P. W. Patil and S. Murala, "MSFgNet: A novel compact End-to-End deep network for moving object detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 11, pp. 4066–4077, Nov. 2019.

[83] X. Ou *et al.*, "Moving object detection method via ResNet-18 with encoder–decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108152–108160, 2019.

[84] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 598–605.

[85] F. Cozman and E. Krotkov, "Depth from scattering," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 801–806.

[86] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 167–179, Feb. 1998.

[87] N. Hautiere, J. P. Tarel, and D. Aubert, "Towards fog-free in-vehicle vision systems through contrast restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[88] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, pp. 325–332.

[89] E. Namer and Y. Y. Schechner, "Advanced visibility improvement based on polarization filtered images," *Proc. SPIE*, vol. 5888, Aug. 2005, Art. no. 588805.

[90] D. Miyazaki, D. Akiyama, M. Baba, R. Furukawa, S. Hiura, and N. Asada, "Polarization-based dehazing using two reference objects," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 852–859.

[91] J. M. Guo, J. Y. Syue, V. R. Radzicki, and H. Lee, "An efficient fusion-based defogging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4217–4228, Sep. 2017.

[92] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.

[93] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.

[94] J. John and M. Wilscy, "Enhancement of weather degraded video sequences using wavelet fusion," in *Proc. 7th IEEE Int. Conf. Cybern. Intell. Syst.*, Sep. 2008, pp. 1–6.

[95] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[96] B. Xie, F. Guo, and Z. Cai, "Improved single image dehazing using dark channel prior and multi-scale retinex," in *Proc. Int. Conf. Intell. Syst. Design Eng. Appl.*, Oct. 2010, pp. 848–851.

[97] S. C. Pei and T. Y. Lee, "Nighttime haze removal using color transfer pre-processing and dark channel prior," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 957–960.

[98] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 617–624.

[99] Y. Gao, L. Yun, J. Shi, F. Chen, and L. Lei, "Enhancement MSRCR algorithm of color fog image based on the adaptive scale," *Proc. SPIE*, vol. 9159, Apr. 2014, Art. no. 91591B.

[100] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4572–4576.

[101] Z. U. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *Proc. SPIE*, vol. 13, no. 1, pp. 100–111, 2004.

[102] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[103] B. Xie, F. Guo, and Z. Cai, "Universal strategy for surveillance video defogging," *Opt. Eng.*, vol. 51, no. 10, 2012, Art. no. 101703.

[104] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1701–1708.

[105] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, Jul. 2012.

[106] N. S. Pal, S. Lal, and K. Shinghal, "A robust framework for visibility enhancement of foggy images," *Eng. Sci. Technol., Int. J.*, vol. 22, no. 1, pp. 22–32, 2019.

[107] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[108] C. Xiao and J. Gan, "Fast image dehazing using guided joint bilateral filter," *Vis. Comput.*, vol. 28, nos. 6–8, pp. 713–721, Jun. 2012.

[109] J. P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2201–2208.

[110] G. Yadav, S. Maheshwari, and A. Agarwal, "Foggy image enhancement using contrast limited adaptive histogram equalization of digitally filtered image: Performance improvement," in *Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Sep. 2014, pp. 2225–2231.

[111] J. Y. Kim, L. S. Kim, and S. H. Hwang, "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 475–484, Apr. 2001.

[112] T. K. Kim, J. K. Paik, and B. S. Kang, "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering," *IEEE Trans. Consum. Electron.*, vol. 44, no. 1, pp. 82–87, Feb. 1998.

[113] Y. T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.

[114] Q. Wang and R. Ward, "Fast Image/Video contrast enhancement based on weighted thresholded histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 757–764, May 2007.

[115] S. Salazar-Colores, I. C. Aceves, and J. M. R. Arreguin, "Single image dehazing using a multilayer perceptron," *Proc. SPIE*, vol. 27, no. 4, 2018, Art. no. 043022.

[116] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2995–3000.

[117] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 154–169.

[118] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-net: All-in-One dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4770–4778.
[119] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An End-to-End system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
[120] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1548–1560, Jun. 2018.
[121] H. Zhang, V. Sindagi, and V. M. Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 902–911.
[122] Y. Li, F. Guo, R. T. Tan, and M. S. Brown, "A contrast enhancement framework with JPEG artifacts suppression," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 174–188.
[123] S. Bhattacharya, S. Gupta, and K. S. Venkatesh, "Dehazing of color image using stochastic enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2251–2255.
[124] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.
[125] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.
[126] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 576–591.
[127] Y. S. Lai, Y. L. Chen, and C. T. Hsu, "Single image dehazing with optimal transmission map," in *Proc. 21st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 388–391.
[128] Y. Bahat and M. Irani, "Blind dehazing using internal patch recurrence," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2016, pp. 1–9.
[129] J. He, C. Zhang, R. Yang, and K. Zhu, "Convex optimization for fast image dehazing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2246–2250.
[130] J. H. Kim, W. D. Jang, J. Y. Sim, and C. S. Kim, "Optimized contrast enhancement for real-time image and video dehazing," *J. Vis. Commun. Image Represent.*, vol. 24, no. 3, pp. 410–425, Apr. 2013.
[131] X. Fan, Y. Wang, X. Tang, R. Gao, and Z. Luo, "Two-layer Gaussian process regression with example selection for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2505–2517, Dec. 2017.
[132] Y. Xu, J. Wen, L. Fei, and Z. Zhang, "Review of video and image defogging algorithms and related studies on image restoration and enhancement," *IEEE Access*, vol. 4, pp. 165–188, 2016.

Sourav Dey Roy (Student Member, IEEE) received the B.E. and M.Tech. degrees in computer science and engineering from Tripura University (A Central University), Suryamaninagar, Tripura, India, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree in computer science and engineering, under the sole supervision of Mrinal Kanti Bhowmik. He has worked as a Junior Research Fellow (JRF) under a Defence Research and Development Organization (DRDO) funded project, Government of India, from June 2018 to April 2019. He is currently working as a Senior Research Fellow (SRF), sponsored by the Council of Scientific and Industrial Research (CSIR), Government of India. His research interests include computer vision, object detection, face and facial expression recognition, medical imaging, and deep learning.

Mrinal Kanti Bhowmik (Senior Member, IEEE) received the B.E. degree in computer science and engineering from the Tripura Engineering College in 2004, the M.Tech. degree in computer science and engineering from Tripura University (A Central University), India, in 2007, and the Ph.D. degree in engineering from Jadavpur University, Kolkata, India, in 2014. He has successfully completed two Department of Electronics and Information Technology (DeitY) (Now Ministry of Electronics and Information Technology (MeitY)) funded projects, one the Department of Biotechnology (DBT)-Twinning project, one Society for Applied Microwave Electronics Engineering and Research (SAMEER) funded project, and one Indian Council of Medical Research (ICMR) project as the Principal Investigator. He is currently the Principal Investigator of a Defense Research and Development Organization (DRDO), Government of India funded project. Since July 2010, he has been serving with the Department of Computer Science and Engineering, Tripura University as an Assistant Professor. He was awarded the Short Term Indian Council of Medical Research (ICMR), Department of Health Research (DHR) International Fellowship from 2019 to 2020 as a Senior Indian Biomedical Scientist for bilateral cooperation in cross-disciplinary research area (i.e., biomedical diagnostic and inferencing systems). His current research interests are in the field of computer vision, security and surveillance, medical imaging, and biometrics.