TU-VDN: TRIPURA UNIVERSITY VIDEO DATASET AT NIGHT TIME IN DEGRADED ATMOSPHERIC OUTDOOR CONDITIONS FOR MOVING OBJECT DETECTION

Anu Singha, Mrinal Kanti Bhowmik
Department of Computer Science and Engineering, Tripura University (A Central University), Agartala, India
anusingh5012@gmail.com, mrinalkantibhowmik@tripurauniv.in

ABSTRACT

Even though thermal infrared images captured during night time are available in some publicly available datasets, such images acquisitioned in adverse weather conditions such as low light, dust, rain, fog etc. are not reported as yet to the best of our knowledge. Because of these deficiencies, object detection techniques applicable in weather affected night thermal infrared images have a very limited reporting in literature. In the present scope, we discussed the acquisition, creation, design, and ground truth annotation of a new video dataset consisting of nearly 60 videos representing 4 atmospheric conditions: low light, dust, rain, fog, named as Tripura University Video Dataset at Night time (TU-VDN) in adverse weather conditions, suitable for this purpose. The objective is to provide a night video dataset containing moving objects with annotated ground truth in the image frame sequences. Using TU-VDN a comparative study is made between the results of ten existing state-of-the-art moving object segmentation methods.

Index Terms—Moving Object Detection; Night Time; Thermal Infrared; Tripura University Video Dataset at Night time (TU-VDN); Atmospheric Conditions; Ground Truth.

1. INTRODUCTION

Most automatic night vision systems for monitoring intelligently of moving objects presume that the input images have clear visibility under lane light but unfortunately this does not ensue all the time [1]. The moving object monitoring performance depends closely on the enhanced quality of the images [2]. The quality of outdoor images is affected by several atmospheric conditions that alter the key characteristics (e.g., intensity, colour, polarization, and coherence) of the light source due to scattering by medium aerosols [3, 4]. Due to poor atmospheric conditions, the contrast of the images is degraded, which affects the visibility in such a scenario. The contrast degradation depends on the coefficient of light scattering through aerosols that are suspended in the atmosphere. However, as the atmospheric aerosol size decreases, both the type and amount of scattering change. Smaller aerosols cause more scattering, especially backscattering, and the loss of contrast is more severe [3]. In the last few decades, large datasets have been designed to meet the increasing demands for the development of new models for object detection under poor atmospheric conditions [5, 6]. However, there is still a lack of video datasets for moving object detection tasks that provide balanced coverage in atmosphere-degraded outdoor scenes, especially at night.

Furthermore, for detecting moving objects, both a visual digital camera and a typical charge-coupled device (CCD) camera have the advantage of high resolution, which renders them more suitable for day time or night time use with a proper lighting setup. However, they are ineffective in environments with poor illumination or visibility due to atmospheric conditions because the appearance of objects in the captured images is not as clear as in images that are captured during under normal atmospheric conditions [7, 1]. Several related works have been conducted in such environments [8, 9, 10]. To address the limitations of visual and CCD cameras at night time, many studies have been conducted on methods that detect objects with infrared based cameras [11, 12, 13, 14, 15] – including near-infrared (NIR) and far-infrared (FIR) cameras. NIR cameras are robust against darkness, and they are less costly than FIR cameras. However, NIR cameras have a similar drawback to that faced by CCD cameras when the interferences are produced by vehicle headlights. In addition, the attenuation of visual, CCD, and NIR radiation that is produced through atmospheric aerosols is mostly due to their short wavelengths. In contrast, FIR cameras enable robust object detection regardless of the atmospheric conditions because as the spectrum wavelength increases, the effect of bad atmospheric conditions decreases [4]. Far less research has been carried out on moving object detection at night time under various atmospheric conditions using thermal images because of the high price of FIR cameras.

The contributions of this paper are summarized below:

1. The paper provides the research community with a comprehensive thermal video dataset of outdoor night scenes degraded by different adverse weather conditions like fog, dust, rain, and low light/poor illumination, referred as *Tripura University Video Dataset at Night time (TU-VDN)*.

2. The paper provides annotated ground truth images of the prominent objects in each of the extracted frames of the created video dataset.

3. The paper also provides a comparison of ten most widely used state of-the-art moving object detection methods based on segmentation of foreground/ background and thus, helps to select the most effective detection methods in the weather degraded outdoor scenes.

Rest of the paper is organized as follows: Section 2 briefly describes the review of related datasets and Section 3 describes the design issues and statistics of the created video dataset under different atmospheric conditions at night. In Section 4, the generation of ground truth images of the salient moving objects in each of the extracted frames is described. In Section 5, popular and widely used state-of-the-art object detection techniques are implemented and report the experimental results of these methods on our dataset. And finally, Section 6 concludes the paper.

2. RELATED DATASETS

Frame based object detection is interrelated with video based object detection, background subtraction, and moving object segmentation. In this section, we will review the most related datasets including thermal and visual-thermal since there is no particular dataset available for purely night based or bad atmosphere based such as dust, fog, rain etc.

OSU-T [5]: OSU thermal pedestrian database is a part of OTCBVS benchmark dataset collection for evaluating state-of-art computer vision algorithms. It contains only about person detection in outdoor environment under bad weather of light rain, cloudy, and Haze. This database has captured only 284 numbers of frames from 10 video sequences in day time. The persons present in a frame are annotated by bounding box.

BU-TIV [16]: Thermal Infrared Video (TIV) is the only dataset which has addresses several visual analysis tasks such as single or multi-view object tracking, counting, and group motion etc. It contains 16 video clips with 63,782 frames in total about pedestrian, runner car, bicycle, motorcycle, and bat. Bounding box based annotation has done.

ASL-TID [17]: Thermal Infrared Dataset (TID) is designed for object detection, not for tracking with key challenges, for e.g., moving camera, cluttered background, and occlusion. It contains 4,381 frames divided into 8 sequences about humans, cat, and horse those are manually annotated bounding box based ground truth.

LTIR [18]: The Linköping Thermal InfraRed (LTIR) is a short term single object tracking dataset of 20 video sequences and 11,269 frames. The video clips are recorded cluttered background, occlusion, and size-change objects by static, hand-held, moving camera in indoor as well as outdoor scenes. It captured several objects like rhinoceros, human, horse, car, dog, quadrocopter.

LITIV [19, 20]: This dataset consists of 3 major parts, one of these 'thermal-visible registration' for people tracking. Indoor environment videos are of total 9 sequences of 6,236 frames, and annotations are based on polygons.

AIC-TV [21]: The dataset designed for tracking objects such as people, bicycles, and vehicles. It fused information from standard CCTV and thermal infrared spectrum videos with key challenges like scale variation, dark night-time, and occlusion. It is a small indoor and outdoor dataset of 6 sequences with 2,013 frames in total.

OSU-CT [22]: It is also a part of OTCBVS benchmark dataset, fusion based object detection in color and thermal imagery. The OSU-CT database is about pedestrian only, and consists of 17,089 frames from 6 outdoor video clips.

CVC-14 [23]: A recent pedestrian detection dataset of visible-FIR day-night sequences. It composed by 2 sets of high quality outdoor sequences: the day and night sets. The ground truth annotations have done through bounding box procedure.

KAIST [24]: One of the largest pedestrian dataset consists of 95,000 color-thermal pair frames. To aligned multispectral (i.e. RGB color and thermal) frames, a beam splitter hardware used for various regular traffic scenes at day and night time. All the pairs are manually annotated includes temporal correspondence between bounding boxes.

CDnet 2012 [25]: It is consists of diverse set videos which are covered various change detection challenges in indoor and outdoor scenarios: dynamic background, camera jitter, intermittent object motion, shadows and thermal. The dataset consists of 31 video sequences divided into 6 video categories. And all sequences are accompanied by accurate pixel wise labeling ground truth segmentation.

CDnet 2014 [6]: The 2014 version of CDnet is a extension of CDnet 2012 with additional following change detection challenges: low frame rate, bad weather, night sequences, PTZ, and air turbulence. And a total of 53 video sequences with nearly 1,60,000 frames.

Several of these datasets have been designed in the past to evaluate moving object detection methods. Among these datasets, four have been recorded with thermal sensors to detect and tracking objects (i.e. OSU-T, BU-TIV, LTIR, ASL-TID), where BU-TIV dataset is primarily designed for visual analysis tasks. These datasets only contain day-time video sequences; where OSU-T dataset includes weather conditions with low resolution thermal camera to detect only pedestrian.

Next numerous datasets (LITIV, AIC-TV, OSU-CT, CVC-14, KAIST, CDNet 2012, CDNet 2014) contain both colour and thermal video sequences, few of them (i.e. LITIV, OSU-CT, KAIST) works on fusion between two modalities to robust detection. The night video sequences contains in AIC-TV, CV-14, KAIST, and CDNet 2014. These datasets consist of various challenges, but very rare datasets consider weather conditions except CDNet 2014 although it is day time. As a consequence, it is difficult to evaluate robustness of object detection methods in atmospheric conditions especially in night vision because more than half of object related accidents occur in the night time.

Therefore, we have been designing an atmospheric weather degraded conditions based standard video dataset at night time that cover many real-world scenarios. The considered atmospheric conditions are dust, fog, rain, and a low light environment to take the advantages of the thermal camera.

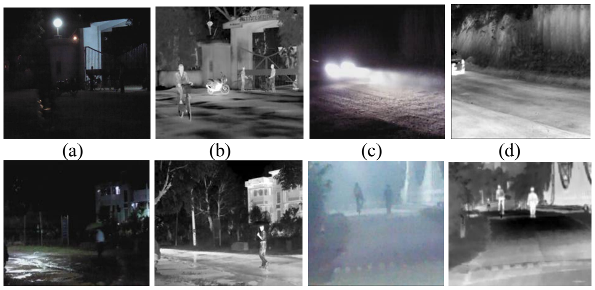
3. Designing Issues And Statistics of Created Dataset

Atmospheric aerosols reduce the visibility of the targets in a scene. This effect is especially debilitating at night. It directly affects the visibility through the aerosols and through vehicle headlamps and, street headlamps. At night, an object is typically visible when light from a source is reflected by the object back to the terminal camera sensors. To detect the presence of objects, terminal sensors use several electromagnetic (EM) spectra that range from the visible to the near-infrared to the far-infrared regions. For electro-optical (EO) sensors, when an EM wave propagates through the atmosphere, the primary factors that are responsible for extinction are *absorption* and *scattering* by atmospheric aerosols (for example, -rain, dust, and fog). Both factors degrade the performance of all sensors [4]. Because the particle size well exceeds the wavelength in the visible portion of the EM spectrum (0.4 to 0.74 μm), attenuation by atmospheric aerosols is independent of the wavelength. Hence, the attenuation is most severe in the visible wavelength range. As the wavelength increases, attenuation becomes less of an issue. Since wavelengths in the far-infrared region exceed those of other infrared wave bands, impact of particles on far-infrared waves is relatively insignificant. Far-infrared waves provide the advantage of 'seeing' not only at night but also through many atmospheric aerosols such as dust, fog, and rain. Fig. 1 shows, visual frames and the corresponding thermal sample frames that were captured at night under several atmospheric conditions. The ability to see under low light and through atmospheric particles is useful for security and surveillance applications, which can benefit from the power of thermal imaging.

3.1. Video Recording Conditions and Acquisition Setup

The video sequences in outdoor environment are mainly influenced by several factors, for example - atmosphere, low light at night. Such conditions amend the key characteristics of EM radio wave due to attenuation by atmospheric aerosols [3]. There are several factors (i.e. temperature, dew point, relative humidity, wind speed, weather, visibility and so on) considered during data acquisition so as to reduce the negative influences in analysis.

Temperature, dew point, and relative humidity: The infrared models primarily used temperature and dew point temperature to compute relative humidity. As temperature decreases at night time, the relative humidity increases even though amount of water vapour in the air remains same. In night the temperature of the air cools down and often reaches it dew



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 1. Sample Frames of Created Dataset in Night time (a), (b) Visual and Corresponding Thermal Frame in Low Light Condition respectively; (c), (d) Visual and Corresponding Thermal Frame in Dust Condition respectively; (e), (f) Visual and Corresponding Thermal Frame in Rain Condition respectively; (g), (h) Visual and Corresponding Thermal Frame in Fog Condition respectively.

point temperature, so the water vapour in the air changes to visible liquid droplets. In other words, fog droplet form when the difference between air temperature and dew point is less than 2.5^0C with relative humidity approximately 100% [26]. General FIR sensor performance can be predicted to a certain degree by knowing the humidity value in the area of interest i.e. the greater the humidity, the greater the amount of water vapour present, and the greater the infrared (IR) absorption. In general, the range which has considered during the data acquisition to capture the influence of climate in IR sensor is – temperature 2^0C-30^0C, humidity 80%-100%, dew point 2^0C-15^0C over whole year for capturing seasonal data.

Wind speed: Winds may affect the dust area of interest for all EO sensors by increasing the density of particles in the air, which can impact FIR performance because it also decreases thermal contrast that in turn reduces the sensors image quality. It is important to accurately forecast wind speed; we managed range from 1 mph to 4.5 mph for our acquisition. In this consequence, the dust particles which basically comes from soil lifted by wind, we captured it from vehicles travelling areas on under construction roads.

Precipitation: Raindrops begin forming when water vapour condenses on micro-meter sized particles floating in the atmosphere. These particles grow to millimetre sized droplets, which are heavy enough to begin falling based upon precipitation rate. Here, we consider rate of precipitation in between 2.5 mm (0.098 inch) – 7.6 mm (0.30 inch) or 10 mm (0.39 inch) per hour to capture the moderate rain falls.

3.2. Dataset Features

The TU-NVD dataset provides realistic diverse set of outdoor videos in night vision over thermal modality. By maintaining above mentioned acquisition conditions, the current dataset consists of total 60 video sequences under different atmospheric conditions. Each video clip of 2 minutes duration is recorded with FLIR camera rigidly mounted 90^0 alignments on a tripod stand by maintaining 200M to 2KM distance from objects. Conversely for motion background, the video is captured by mounting the camera on a moving vehicle (20~30 km/h) where the objects, camera, and background are moving

(a) (b) (c)

Figure 2. A sample frame from created dataset for ground truth generation (a) a thermal frame in dust condition (b) identified moving (√) and non-moving (×) objects in the frame (c) corresponding ground truth binary mask of the frame respectively. **1 frame is annotated per 10 frames.

simultaneously. Overall statistics has listed in Table 1. The key features of the designed dataset are as follows:

(i) Each frame contains multiple types of moving objects, e.g., pedestrians, various types of vehicles, bicyclists, motorbikes, trains, and pets.

(ii) The night video clips were captured under three outdoor atmospheric scenarios, namely, dust, rain, and fog, which produce flat regions in thermal scenes. In addition, the captured scenes are mostly in urban areas, which correspond to larger surface variations due to the presence of hot and cool objects such as houses, warehouses, office buildings, streets, and residents. Therefore, areas with varied background and adverse weather conditions produce thermal characteristics that lead to an increased *flat cluttered* region in the target area.

(iii) A conventional challenge is encountered, namely, *a dynamic background* due to shaking trees, since the whole dataset was recorded in an outdoor environment.

(iv) The key issue with the FIR camera is *thermal temperature adjustment* during the maiden appearance of a moving object in a video sequence, which causes illumination-type effects in the background model from the current video frame.

(v) Motion-camera-based videos are captured by mounting the camera on a moving vehicle, where the camera and objects are moving and shaking simultaneously.

4. Ground Truth Generation of Moving Salient Objects on the Created Dataset

To test the efficiency of object detection algorithms, the ground truth generation of targets in a video sequence is very essential. Here we have adopted pixel level binary mask based ground truth to evaluate moving object detection methods. However, manual annotation of an accurate ground truth data often results in uncertainty and strong subjective bias. As well, the visual analysis by group members based annotation is impractical due to man-power and time constraints. And it is also particularly difficult for a person to reliably identify actual number moving objects in a multi object frame because all objects not always move simultaneously, as shown in Fig. 2.

Therefore, we decided to produce pixel level ground truth images for our dataset with the following semi-automatic procedures: (i) To get a good approximation of the targets, one to estimate a background model [27, 28, 29] which will be compared with current image. This automatic mechanism could easy to identify the regions of interest in a frame. (ii) In second stage, a user will supervise the identified regions of interest by some small changes manually to ensure of required quality in labeling work. (iii) As a consequence, only one user

Table 1: Statistics of Created Dataset in different Atmospheric Conditions at Night Time.

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

Table 2: Comparison using average F_1-Score, MCC, and Accuracy performance measures with ten different methods on the TU-NVD dataset. Color's used to indicate place of methods – green for best methods, blue for second best methods, and red for worst methods.

State-of-Art	Low Light			Dust			Rain			Fog		
Methods	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.	F_1-Score	MCC	Acc.
Vibe	0.5738	0.5954	0.9907	0.5565	0.5823	0.9740	0.7307	0.7379	0.9877	0.5075	0.5484	0.9851
Subsense	0.4812	0.5091	0.9837	0.5405	0.5679	0.9722	0.6112	0.6171	0.9788	0.7649	0.7663	0.9947
LOBSTER	0.5244	0.5413	0.9890	0.4967	0.5346	0.9688	0.5658	0.5949	0.9793	0.6225	0.6377	0.9916
PAWCS	0.3155	0.3505	0.9870	0.2322	0.2872	0.9656	0.6628	0.6752	0.9875	0.2224	0.2376	0.9898
PBAS	0.5668	0.5803	0.9901	0.4033	0.4311	0.9547	0.7035	0.6893	0.9823	0.6686	0.6788	0.9929
Multicue	0.4961	0.5314	0.9798	0.6166	0.6345	0.9714	0.5511	0.5912	0.9717	0.4419	0.5160	0.9748
KDE	0.3845	0.3996	0.9691	0.2689	0.2978	0.9453	0.6198	0.6315	0.9606	0.4228	0.4313	0.9897
MoG_V2	0.3119	0.3486	0.9854	0.2208	0.2799	0.9673	0.3532	0.3963	0.9651	0.1928	0.2344	0.9895
Eigenbackground	0.3695	0.4190	0.9673	0.3603	0.3996	0.9184	0.2120	0.1761	0.6499	0.3634	0.4339	0.9745
Codebook	0.3093	0.3623	0.8848	0.2066	0.2245	0.6807	0.2736	0.3958	0.9021	0.1989	0.2943	0.8809

can reliably classify pixels belongings to either static or moving class: **Static-** assigned binary value of 0, **Moving-** assigned binary value of 1.

The whole process has been verified time to time by our research team. In this semiautomatic way, one person can produce rapidly an uncontroversial binary ground truth images for camera captured videos. To validate our annotation results, we used TSLAB annotation tool [30] - an advanced and user friendly tool for fast labeling of moving objects. Finally, we compare the annotated data produced by both procedures (i.e. our semiautomatic and TSLAB) with similarity score of 95% enough to ensure the effectiveness of our annotation.

5. Comparision of Image Sequences Segmentated by the State-of-Art Object Detection Methods

Moving objection segmentation at night time under thermal medium degraded by adverse atmospheric conditions has been one of the major research topics. Numerous computer aided detection techniques have been proposed in the literature for segmentation of video frames. In our work, we have used selected most popular ten methods for comparative study over our *TU-NVD* dataset. These state-of-art object detection methods are Vibe [28], Subsense [29], LOBSTER [31], PAWCS [32], PBAS [27], Multicue [33], KDE [34], MoG_V2 [35], Eigenbackground [36], Codebook [37] respectively. To provide a better assessment of overall performance and compare the performances among state-of-art methods, we used metrics like Accuracy, F_β-score and Matthew's Correlation Coefficient (MCC).

For moving object segmentation evaluation, three videos from three challenges (i.e. flat cluttered background, dynamic background, motion camera) for each of the weather conditions (i.e. low light, dust, rain, and fog) are selected. The average value of these performance metrics over three challenges for each weather degraded thermal image sequences has shown in Table 2.

From this comparative analysis, we have analysed performances over weather conditions. In the *low light condition*, the *Vibe* method is providing better F_1-score, MCC and Accuracy, where *Codebook* and *MoG_V2* shows lowest results. In the *dust condition*, the performance using *Multicue* in all category metrics is satisfactory found as finest method, and 0.25% decreased accuracy than *Vibe* which has revealed as second best method. As usual, *Codebook* shows lowest results hereto. In the *rain condition*, *Vibe* and *PBAS* provides most promising metric values respectively, and *Eigenbackground* method shows awful outcome. In the *fog condition*, the *Subsense* and *PBAS* methods are providing best results, where *MoG_V2* shows lowest results. To provide a better visual understanding about the categorization results, we have shown a typical segmentation results in Fig. 3 for various atmospheric conditions via ViBe background subtraction method.

(a) Low Light (b) Dust (c) Rain (d) Fog

Figure 3. Typical segmentation results for various atmospheric conditions in our created night time dataset. row (1) shows input frames, row (2) shows ground truth, row (3) shows *ViBe* results (best) , row (4) shows *CodeBook* results (worst).

6. Conclusion and Future Work

We have described the acquisition and design setup of a newly created night video dataset TU-VDN for moving object detection on thermal infrared images. The dataset consists of (a) degraded atmospheric night outdoor scenes under low light, dust, rain, and fog; (b) semi-automatic moving object annotations. The far-infrared sensor has shown great video sequence acquisitions under bad atmospheric conditions because of its higher spectrum wavelength. Furthermore, the paper investigates the potentiality of the some well-known moving object detection techniques based on background segmentation. The evaluation metrics demonstrates that *Vibe*, *Subsense*, *PBAS*, and *Multicue* methods are showing superior performances where *MoG_V2*, *Codebook* and *Eigenbackground* have still been worst for almost all conditions. In future, the dataset will be regularly reworked and extended to include other atmospheric conditions. Also we will develop new moving object detection algorithm as well as deep learning based detection to overcome the limitations of the state-of-the-art methods.

Acknowledgement

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

7. References

[1] Tan RT. Visibility in bad weather from a single image. *In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2008. p.1-8.

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

[3] Narasimhan SG, Nayar SK. Vision and the atmosphere. *International Journal of Computer Vision* 2002; **48**: 233-254.

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

[5] Davis j, Keck M. A two-stage approach to person detectionin thermal imagery. *In: Workshop on Applications of Computer Vision (WACV)*; 2005.

[6] Wang Y, Jodoin PM, Porikli F, Konrad J, Benezeth Y, Ishwar P. CDnet 2014: An Expanded Change Detection Benchmark Dataset. *In: IEEE Workshop on Change Detection (CDW)*; 2014. p. 387-394.

[7] Mahlke S, Rösler D, Seifert K, Krems JF, Thüring M. Evaluation of six night vision enhancement systems: Qualitative and quantitative support for intelligent image processing. *Hum. Factors* 2007; **49**: 518–531.

[8] Tsimhoni O, Bärgman J, Flannagan M. Pedestrian detection with near- and far-infrared night vision enhancement. *Leukos* 2007; **4**: 113–128.

[9] Narasimhan SG. Models and algorithms for vision through the atmosphere. Diss. Columbia University; 2003.

[10] Henry RC, Mahadev S, Urquijo S,Chitwood D. Color perception through atmospheric haze. *JOSA A 17.5* 2000; 831-835.

[11] Ge J, Luo Y, Tei G. Real-time pedestrian detection and tracking at nighttime for driver-assistance systems. *IEEE Trans. Intell. Transp. Syst.* 2009; **10**: 283–298.

[12] Han J, Gaszczak A, Maciol R, Barnes SE, Breckon TP. Human pose classification within the context of near-IR imagery tracking. *In: 9th SPIE 8901 Opt. Photon. Counterterrorism, Crime Fight. Def.* 2013; 1–11.

[13] Bertozzi M. IR pedestrian detection for advanced driver assistance systems. *In: 25th DAGM Symp. Pattern Recognit.* 2003; 582–590.

[14] Treptow A, Cielniak G, Duckett T. Real-time people tracking for mobile robots using thermal vision. *Robot. Autonom. Syst.* 2006; **54**: 729–739.

[15] Dai C, Zheng Y, Li X. Pedestrian detection and tracking in infrared imagery using shape and appearance. *Comput. Vis. Image Understand.* 2007; **106**: 288–299.

[16] Wu Z, Fuller N, Theriault D, Betke M. A thermal infrared video benchmark for visual analysis. *In: 10th IEEE Workshop on Perception Beyond the Visible Spectrum (PBVS)*; 2014.

[17] Portmann J, Lynen S, Chli M, Siegwart R. People detection and tracking from aerial thermal views. *In: IEEE International Conference on Robotics and Automation (ICRA);* 2014.

[18] Berg A, Ahlberg J, Felsberg M. A thermal object tracking benchmark. *In: 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*; 2015.

[19] Torabi A. An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications. *Jour. Comp. Vis. and Im. Underst.* 2012; **116**:210 – 221.

[20] St-Charles PL, Bilodeau GA, Bergevin R. Online Multimodal Video Registration Based on Shape Matching. *In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*; 2015.

[21] Conaire C, O'Connor N, Cooke E, Smeaton A. Comparison of fusion methods for thermo-visual surveillance tracking. *In: IEEE Conf. Information Fusion*; 2006.

[22] Davis JW, Sharma V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Jour. Comp. Vis. and Im. Underst.* 2007; **106**:162 – 182.

[23] González A, Fang Z, Socarras Y, Serrat J, Vázquez D, Xu J, ALópez AM. Pedestrian Detection at Day/Nighttime with Visible and FIR Cameras: A Comparison. *Sensors* 2016; **16**: 820.

[24] Hwang S, Park J, Kim N, Choi Y, Kweon IS. Multispectral Pedestrian Detection: Benchmark Dataset and Baseline. *In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015. p. 1037-1044.

[25] Goyette N, Jodoin PM, Porikli F, Konrad J, Ishwar P. Changedetection.net: A new change detection benchmark dataset. *In: IEEE Workshop on Change Detection (CDW) at CVPR, Providence, RI*; 2012.

[26] Kumari A, Thomas PJ, Sahoo SK. Single image fog removal using gamma transformation and median filtering. *In: Annual IEEE India Conference (INDICON)*; 2014.p. 1-5.

[27] Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter. *In: IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*; 2012. p. 38–43.

[28] Barnich O, Droogenbroeck MV. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.* 2011; **20**:1709–1724.

[29] St-Charles PL, Bilodeau GA, Bergevin R. SuBSENSE: A Universal Change Detection Method with Local Adaptive Sensitivity. *IEEE Trans. on Image Processing* 2015; **24**: 359-373.

[30] Cuevas C, Yáñez EM, García N. Tool for Semiautomatic Labeling of Moving Objects in Video Sequences: TSLAB. *Sensors* 2015; **15**: 15159-15178.

[31] St-Charles PL, Bilodeau GA. Improving Background Subtraction using Local Binary Similarity Patterns. *In: IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2014.

[32] St-Charles PL, Bilodeau GA, Bergevin R. Universal Background Subtraction Using Word Consensus Models. *IEEE Trans. on Image Processing* 2016; **25**: 4768-4781.

[33] Unzueta L, Nieto M, Cortes A. Barandiaran J. Otaegui O, Sanchez P. Adaptive Multicue Background Subtraction for Robust Vehicle Counting and Classification. *IEEE Trans. on Intelligent Transportation Systems* 2012; **13**: 527-540.

[34] Elgammal AM, Harwood D, Davis LS. Non-parametric model for background subtraction. *In: 6th European Conf. on Comput. Vis.*; 2000. p. 751–767.

[35] Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. *In: IEEE Conf. Comput. Vis. Pattern Recognit.*; 1999. p. –252.

[36] Oliver NM, Rosario B, Pentland A. A bayesian computer vision system formodeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; **22**:831–843.

[37] Wu M, Peng X. Spatio-temporal context for codebook-based dynamic background subtraction. *Int. J. Electron. Commun.* 2010; **64**:739 – 747.