Moving Object Detection over Wireless Visual Sensor Networks using Spectral Dual Mode Background Subtraction

Ahmed M. AbdelTawab[1]*, M.B. Abdelhalim[2], S.E.D. Habib[3]

Electronics and Communications Department, Faculty of Engineering, Misr University for Science & Technology MUST, Giza, Egypt[1]

College of Computing and Information Technology, Arab Academy of Science and Technology and Maritime Transport, Cairo, Egypt[2]

Electronics and Communications Department, Faculty of Engineering, Cairo University, Giza, Egypt[3]

***Abstract*—Wireless Visual Sensor Networks (WVSN) play an essential role in tracking moving objects. WVSN's key drawbacks are storage, power, and bandwidth. Background subtraction is used in the early stages of target tracking to extract moving targets from video images. Many standard methods of subtracting backgrounds are no longer suitable for embedded devices because they use complex statistical models to manage small changes in lighting. This paper introduces a system based on the Partial Discrete Cosine Transform (PDCT), reducing the vast dimensions of processed data while retaining most of the important information, thereby reducing processing and transmission energy. It also uses a dual-mode single Gaussian model (SGM) for accurate detection of moving objects. The proposed system's performance is to be assessed using the standard CDnet 2014 benchmark dataset in terms of detection accuracy and time complexity. Furthermore, the suggested method is compared to previous WVSN background subtraction methods. Simulation results show that the proposed method consistently has 15% better accuracy and is up to 3 times faster than the state-of-the-art object detection methods for WVSN. Finally, we showed the practicality of the suggested method by simulating it in a sensor network environment using the Contiki OS Cooja Simulator and implementing it in a real testbed using Cortex M3 open nodes of IOT-LAB.**

Keywords*—Background subtraction; discrete cosine transform; embedded camera networks; Gaussian mixture models; wireless visual sensor networks

I. Introduction

Wireless sensor networks (WSNs), which are made up of thousands of scalar sensors nodes that are spatially distributed and wirelessly communicated, have attracted researchers' interest [1]. Small and low-power CMOS cameras and microphones are used in Wireless Visual Sensor Networks (WVSNs), which can collect visual cues from the environment. The WSN's capabilities are being expanded to include sophisticated environmental monitoring, advanced health care delivery, traffic avoidance, fire prevention, and monitoring, as well as object tracking, and modern surveillance systems [2]. WVSN has focused on military, commercial traffic management, and precision agriculture surveillance applications [3]. Three major problems make WVSNs lack vision processing capability. First, sensor nodes' visual processing capability, second, memory storage constraints for sensor nodes and Finlay; communication of large volumes of image data. However, maximising network lifespan while processing huge volumes of multimedia data while following application-specific QoS requirements such as latency, packet loss, bandwidth, and throughput is a challenge. In addition to developing energy-sensitive multimedia processing algorithms and infrastructures, it is also necessary to establish efficient communication strategies [3].

Object detection is the first and most critical step in target tracking [4]. Robust object detection is typically the dominant consumer of processing and resources, where the moving targets are extracted from the video frames to perform further high-level processing. Lighting changes, shifting backgrounds, artificial or fast motion, and occlusion make accurate foreground object segmentation challenging [5]. The major methodologies for completing the object detection task include optical flow [6], frame differencing [7], and background subtraction [8].

Background subtraction is a standard and consistent method for detecting moving foreground that involves subtracting the background model from the current frame and changing the background model on a regular basis to remove the effects of illumination and inappropriate events. This method is extensively used for motion detection tasks in dynamic scenarios. In practice, basic techniques like mixture of Gaussians (MOG) [9], KDE [10], codebook [11], and ViBe [12] are employed for real applications. Despite the accuracy and efficiency of the MoG [9], the evaluation in [13] demonstrates that MoG can only handle three frames per second on the Blackfin DSP camera nodes with a low image resolution frame size of 320 × 240. The need to update the MoG probability distribution parameters accounts for the long computation time of MoG.

This work aims to investigate the development of moving object detection over WVSN. The Discrete Cosine Transform (DCT) [14] is a frequently utilised image compression technique over WVSN [15, 16]. The DCT algorithm converts signals from the spatial domain to a frequency domain representation. We apply the DCT to minimise the dimensionality of the background subtraction problem while

*Corresponding Author.

maintaining accuracy. The following are the contributions made by this paper:

- A new compression-based background subtraction called Spectral Dual Mode Background Subtraction (SDMBS) uses Partial Discrete Cosine Transform (PDCT) [15] (for dimensionality reduction) and Dual mode SGM [17] (for accuracy) to model the background and distinguish the foreground from the background.
- We implement our approach and compare it to MoG and other compressed-based MoG methods to demonstrate the computational efficiency of our suggested methods. According to the results, our method is up to 10 times faster than the original MoG and three times faster than the compressed-based MoG.
- To demonstrate the algorithm's ability to work in wireless sensor network environments, we simulated and realised the proposed SDMBS in a Cooja network simulator and on the IOT-LAB M3 board.

The rest of the paper is organized as follows. We first present the related work in Section II. We then present a detailed account of the proposed SDMBS approach in Section III. Section IV discusses the simulation results and performance evaluation in detail. Section V draws the paper's conclusion.

II. Related Work

A. Object Detection in WVSN

In visual sensor networks, the cost of data communication is usually far higher than the cost of image processing. As a result, traditional object detection methodologies are ineffective for monitoring and surveillance applications; instead, the image raw data is sent to the sink node, where detection methods are used to determine the moving object. Alternative approaches are to either compress the image at the sensor node and apply object detection at the sink node after decompression, or process the frame before transmission and transmit the useful information or features for further analysis at the sink node. Compression can be applied using Compressed Sensing (CS), wavelet, or DCT. In the second approach; frame processing is applied either on raw image data or compressed domain to further reduce processing complexity at the sensor node. The compressed data is already computed and has less storage space than the raw image frame. The two approaches are briefly reviewed in this section.

1) Compressed data: According to Robust Primary Component Analysis (RPCA) [18], DECOLOR [19], the basic concept of low-order factorization structures and sparse factorization is to divide a given matrix of acquired frames into background and sparse foreground by outliers. The goal of Compressed Sensing CS (low-rank BS) [20] is to send a compressed image to the base station using Compressed Sensing (CS) [13] and then use Orthogonal Matching Pursuit (OMP) [21] to rebuild the image at the receiver end. The authors of [22] proposed a CS-based detection approach that uses CS measurements of a moving object to reconstruct the foreground in a video.

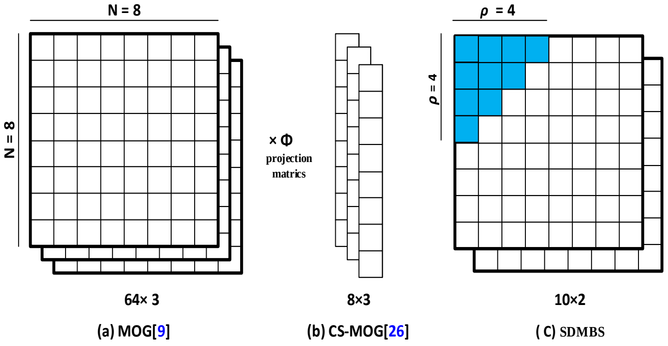
2) Processed data: Because the video to be sent in surveillance applications is generally static, a resource-constrained environment like WVSN does not require the transmission of the entire video. The video can be processed using a compression-based background removal technique to recognise moving objects and send only the foreground data to the monitoring location to save energy and bandwidth. A method for sending image portions instead of the whole image us describes in [23]. It ensures that the sink node receives the bare minimum of image content, as assessed by in-node energy consumption and reconstructed picture peak signal-to-noise ratio (PSNR). The image processing block (Running Gaussian Average technique for object extraction and DWT for ROI transmission) operates at a high frequency to facilitate rapid processing and is only engaged by a separate network processor when images need to be analysed [24]. Because it runs continually, the network processor block is designed to operate at a low frequency. The suggested approach for image processing and communication requires relatively little energy, as evidenced by practical test and simulation results. To save transmission energy, Nandhini et al. [25] propose a method for detecting objects with fewer measures that combines a mean measurement differencing approach with an adaptive threshold strategy.

CS-based background subtraction is measured based on the node before object information is sent, reducing complexity in terms of power, storage, and bandwidth. CSMOG [26] applies MOG [9] to low-rate CS measurements. CSMOG [26] is based on the idea of reducing the number of dimensions in data while still capturing the majority of the information via a random projection matrix. The CSMOG method is consistently superior, up to 6x faster, and uses significantly fewer resources than the standard method, according to real-time requirements. The DWT-based CS object identification framework [27] uses a simple measurement matrix termed the deadweight tonnage block diagonal matrix to refine the pixel-based foreground following the block-based foreground recognition phase in the first stage. The averaging approach using the Adaptive Threshold Technology (MMDATS) in [25] is based on the framework for robust subspace learning. The OMP approach is used to reconstruct the object from foreground measurements. Due to its excellent directional selectivity and shift-invariance, [28] uses a motion segmentation algorithm based on interframe differentiation using the complex Daubechies wavelet transform in the wavelet compression domain.

To reduce the storage space and time required, a background statistical subtraction approach [29] based on motion segmentation in the compression transform domain using Wavelet has been proposed. A good observation was made in 8x8 blocks using the DCT coefficients of the pre-coded JPEG image [22]. They developed a background subtraction strategy that properly depicts the background model over time using competing Hidden Markov Models (HMM). Three techniques for modelling the background directly from the compressed video are presented in [30].

Moving average, median, Gauss blending. These methods use the DCT coefficient (including the AC coefficient) to characterise the background at the block level, and then update the DCT coefficient to match the background. Popa et al. [31] use low DCT compressed area processing to simulate the background. Processing at the block level instead of the pixel level reduces the number of simulation parameters by almost one-third. They also reduce the number of coefficients per block from 64 to 16 while retaining segmentation quality. In the DCT domain, Ye et al. [32] evaluated the background stability and separability of objects. The suggested method restores the target by suppressing the background coefficients by modelling the background as a single Gaussian model for each frequency point. A quaternion-DCT for infrared target recognition is presented by [33]. This approach shows how to create a quaternion with two-directional features (motion feature and kurtosis feature). The QDCT drawing feature acts as a unique signature that helps solve problems when finding small targets. To reduce complexity and simplify hardware implementation, Manimozhi et al. [34] employed a diagonal matrix of binary substitution blocks as the measurement matrix for both DCT-based and DWT-based CS procedures.

According to related research, a large volume of video is required, as well as a significant amount of storage space and processing time for the segmentation method. Compression-based processing is recommended for restricted WVSNs to address the above issues. As a result, we'll describe a motion segmentation method using the DCT in the compression transform domain based on statistical background subtraction. The dual-mode SGM-based background subtraction technique recognises just the foreground blocks of the discrete cosine transform's detailed component to reduce processing complexity. Then, adjust the foreground block to recognise the foreground object. The foreground block is moved to the sink side for rebuilding and tracking. In Fig. 1, the proposed SDMBS (page size = 4) is compared to the original MOG [9] and with block measurements based on compressed sensing (CSMOG) [26].



(a) MOG[9] (b) CS-MOG[26] (C) SDMBS

Fig. 1. Block Computation for (a) MOG, (b) CS-MOG and (c) SDMBS.

III. Proposed System Model

We first describe the steps of Spectral Dual-Mode Background Subtraction (SDMBS), then justify the use of dual-mode SGM (D-SGM) on top of the reduced dimension data PDCT. Fig. 2 depicts the proposed SDMBS algorithm's block diagram as well as the network topology.

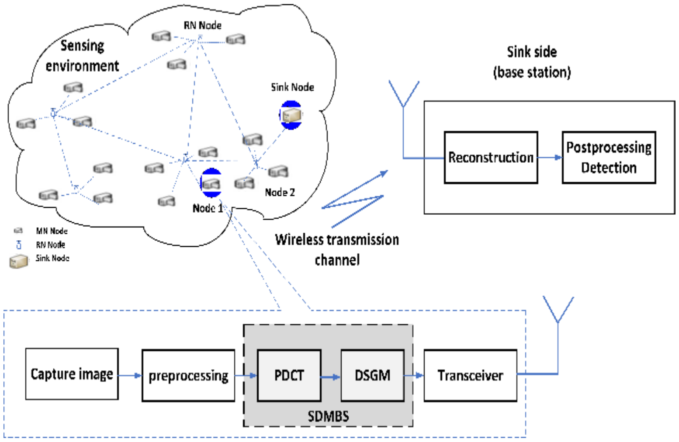


Fig. 2. The Proposed Block Diagram for WVSN-based Object Detection.

A. Network Model

We are considering randomly deploying WVSN nodes in the surveillance field. Each WVSN node is constrained in terms of process and memory resources. The WVSN system model is composed of N visual sensor nodes, Relaying Nodes (RNs), one or more Monitoring Node or Sensor node (MNs), and a Sink Node (SN) [23]. Each sensor node i is thought to be in a 'wakeup' state according to a unique duty cycle $\beta_s \in [0,1]$ during a period t_s to successfully send an image via the network. Thus, it avoids any conflicts induced by two or more nodes simultaneously broadcasting image data. Thus, each sensor is awake for a length of time $\beta_s t_s$ and sleeps for a length of time $(1-\beta_s)t_s$. The frame count is set to zero when a sensor node enters a 'wake up' condition.

B. Pre-Processing

Simple spatial Gaussian filtering and median filtering are used to suppress salt and pepper and Gaussian noise in images captured during the preprocessing step [27]. The filtered frame is then divided into equal-sized blocks, with the SDMBS algorithm applied to each block separately. This can be done in parallel, further reducing computation time.

C. Discrete Cosine Transform (PDCT)

As seen in Fig. 2, each video frame is subsequently divided into 8 × 8 blocks. After that, each block is subjected to DCT. Each 8 × 8 DCT block is represented by the first ten low-frequency DCT components. The partial DCT has the advantage of compressing an 8 × 8 block into 10 samples, which is useful for WSNs with limited resources. Although the rest of the data is sparse, the DCT DC-coefficient stays concentrated in the series' upper left corner. Compressed sensing CS [25] requires a sparse value.

D. Dual Mode Signal Gaussian Model (DM-SGM)

To deal with the inaccuracies that come from modelling the scene using SGMs [35], a dual-mode SGM with age [17] is utilised. While still learning the background reliably, this model safeguards the background model from foreground and noise contamination. The compressed domain PDCT low frequency components are subjected to DM-SGM to identify whether or not the image block contains a moving target. Here, the Gaussian parameter for each grid is computed. Mean, variance, and age are then used to model the background, determining and updating the foreground blocks. There are two models for each block; appearance background models and candidate background models. The candidate background model is ineffective until its age exceeds that of the apparent background model. This dual-mode SGM differs from two-version Gaussian combination models (GMM) [9] in that, with a bi-modal GMM, the foreground facts could still contaminate the history. However, with the dual-mode SGM approach, this is no longer the case.

The two models are switched at this point. At the end, the foreground blocks are determined and applied to the pixel refining stage to detect the pixels containing the target within the foreground block, according to the flowchart in Fig. 3.

The group of pixels in grid i at time t is denoted as $\boldsymbol{\mathcal{G}}_{\boldsymbol{i}}^{(\boldsymbol{t})}$, the number of pixels in $\boldsymbol{\mathcal{G}}_{\boldsymbol{i}}^{(\boldsymbol{t})}$ as $\left|\boldsymbol{\mathcal{G}}_{\boldsymbol{i}}^{(\boldsymbol{t})}\right|$, and the observed pixel intensity of a pixel j at time t as $I_j^{(t)}$, and the mean $\mu_i^{(t)}$, the variance $\sigma_i^{(t)}$, and the age $\alpha_i^{(t)}$ of the SGM model applied to $\boldsymbol{\mathcal{G}}_{\boldsymbol{i}}^{(\boldsymbol{t})}$ is updated as

$$\mu_i^{(t)} = \frac{\tilde{\alpha}_i^{(t-1)}}{\tilde{\alpha}_i^{(t-1)}+1}\tilde{\mu}_i^{(t-1)} + \frac{1}{\tilde{\alpha}_i^{(t-1)}+1}M_i^{(t)} \tag{1}$$

$$\sigma_i^{(t)} = \frac{\tilde{\alpha}_i^{(t-1)}}{\tilde{\alpha}_i^{(t-1)}+1}\tilde{\sigma}_i^{(t-1)} + \frac{1}{\tilde{\alpha}_i^{(t-1)}+1}V_i^{(t)} \tag{2}$$

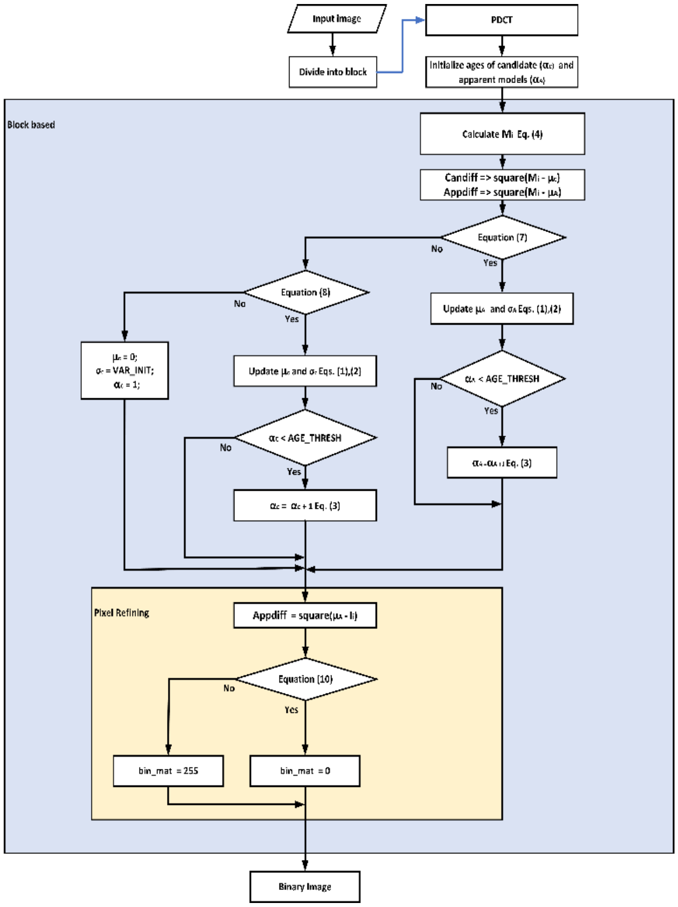


Fig. 3. A Flowchart for the DSGM Process.

$$\alpha_i^{(t)} = \tilde{\alpha}_i^{(t-1)} + 1 \tag{3}$$

$$M_i^{(t)} = \frac{1}{|\mathrm{G}_i|}\sum_{j\in\mathrm{G}_i} I_j^{(t)} \tag{4}$$

$$V_i^{(t)} = \max_{j\in\mathrm{G}_i}\left(\mu_i^{(t)} - I_j^{(t)}\right)^2 \tag{5}$$

$$V_i^{(t)} = \left(\mu_i^{(t)} - M_i^{(t)}\right)^2 \tag{6}$$

DM-SGM [17] uses another SGM as a prospective background model. At this point, the candidate background model is rendered ineffectual until it reaches the same age as the apparent background model, at which point the two models are exchanged. We update the mean, variance, and age of the candidate background model and the apparent background model at time t for grid i, $\mu_{C,i}^{(t)}$, $\sigma_{C,i}^{(t)}$, and $\alpha_{C,i}^{(t)}$and $\mu_{A,i}^{(t)}$, $\sigma_{A,i}^{(t)}$, and $\alpha_{A,i}^{(t)}$, respectively, according to (1), (2), and (3), if

$$\left(M_i^{(t)} - \mu_{A,i}^{(t)}\right)^2 < \theta_s\sigma_{A,i}^{(t)} \tag{7}$$

Where $M_i^{(t)}$is the observed mean and θ_s is a threshold parameter. Also, we update $\mu_{C,i}^{(t)}$, $\sigma_{C,i}^{(t)}$, and $\alpha_{C,i}^{(t)}$, according to (1), (2), and (3).

If condition (7) is violated, and if the observed mean matches the candidate background model, then

$$\left(M_i^{(t)} - \mu_{C,i}^{(t)}\right)^2 < \theta_s\sigma_{C,i}^{(t)} \tag{8}$$

If none of the conditions hold, we start the candidate background model with the current observation. Only one of the two models is altered when this process is used, while the other is left alone. If the candidate's age exceeds the apparent meaning, the two backdrop models for the grid are swapped after updating.

$$\alpha_{C,i}^{(t)} > \alpha_{A,i}^{(t)} \tag{9}$$

Once the candidate is exchanged, the background model is initialised. Finally, an apparent background model is solely employed to determine foreground pixels, as stated in Section E. preventing the background model from being distorted by the foreground data that represents the object.

The candidate background model, rather than the apparent background model, learns the foreground data in the dual-mode SGM, preventing the background model from being distorted by the foreground data that represents the moving object in the frame. So, the models are changed and the correct background model is chosen if the candidate background model's age is greater than the apparent background models.

E. Pixels Refining

A foreground block contains both foreground and background pixels. Each video frame contains a large number of background blocks. As a result, we just need to focus on the small number of foreground blocks. To detect which pixels in a foreground block are indeed foreground, a basic background learning technique for each block is created. If we classify a pixel j in a group i as a foreground pixel,

$$\left(I_j^{(t)} - \mu_{A,i}^{(t)}\right)^2 < \theta_d \sigma_{A,i}^{(t)} \quad (10)$$

where θ_d is a threshold parameter, So, instead of the apparent background model learning the foreground data, the candidate background model learns it. Additionally, the correct background model will be chosen if the candidate background model's age is greater than the apparent background model's, where the models will be swapped. As a result, we don't have to be concerned about the model learning inaccurate foregrounds.

F. Computation Complexity

The quantity of elements processed in every frame determines the difficulty of the computation. We can only evaluate the computing complexity of one block because each frame is divided into equal-sized blocks of size 8 × 8 pixels. Because each frame is broken into blocks of 8 × 8 pixels of similar size, we may calculate the computing cost of a single block.

- For the CS process, we consider the original MoG [9], where each pixel is modelled by 3 Gaussians, which means that we need 64 × 3 Gaussians per block.
- For CS-MoG [26], where each projection value requires three Gaussians, the number of Gaussians required per block is 8 × 3 (a factor of 8 reduction).
- For our proposed method, each block is modelled by 2 Gaussian DM-SGM and we proceed over the 10 low-frequency DCT components, which require 10 × 2 number of Gaussians for each block, a reduction by a factor of 9.6 and 1.2 per block w.r.t. the original MoG and CS-MoG, respectively. Experiments show that it is 2.5 times faster than CS-MOG in processing time.

G. Scene Reconstruction

When an image arrives at the sink node, it is superimposed on the previously received reference frame. Because the suggested technique only communicates a fraction of the entire image, the pixel coordinates at the MN node stay unchanged. This allows a portion of the transmitted image to be used to replace pixels in the reference image at the sink node more efficiently. The pixel values are, however, subject to channel distortion due to the transmission environment.

IV. Experimental Results and Analysis

In this section, experimentation and performance evaluation are done to determine the relevance of our proposed method. The experimental dataset and setup are explained, then the qualitative analysis is shown to illustrate the performance of our system, and evaluation for quantitative and execution performance is done to test the accuracy and running time. In addition, the algorithm is also simulated in a sensor network environment using the Cooja Simulator of Contiki OS [36, 37] and realised in a real testbed using IOT-LAB [38].

A. Dataset and Setup

We will present the results of our compressed domain-based moving object detection technique on a standard benchmark dataset, CDnet 2014[1] [39], to demonstrate its effectiveness. The CDnet 2014 data set is divided into 11 categories with different challenges, each of which contains four to six video sequences. Each video sequence consists of 600 to 7999 frames, with resolutions ranging from 320×240 to 720×576. The simulations were run on an Intel Dual Core i7 3.6GHz processor with 8GB of RAM. The code is written in the C++ language. The total number of frame sequences in each dataset was averaged during the experiments.

B. Qualitative Analysis

Fig. 4 and 5 show the results of our moving object detection technique, Spectral Dual-Mode Background Subtraction (SDMBS). Fig. 4 exhibits performance for some of the representative frames from CDnet 's different categories to show performance against all the CDnet 2014 challenges. Fig. 4 and 5 demonstrate the ground truth and detected object discoveries from the original video frames. Comparing the resulting foreground mask to the relevant ground truth demonstrates the robustness of our suggested strategy for detecting moving objects across different categories.

Most of the CDnet 2014 challenges have excellent qualitative performance; nevertheless, the PTZ and camera jitter categories, as shown in Fig. 4, have poor qualitative performance. The worst performance Due to the zooming and moving features of this category, a compensation stage is required before the object detection stage to compensate for the frame movement. Because of the ghosting artefacts created in the videos in this category, the Intermittent Object Motion category is noisy. Background items moving away, abandoned objects, and objects stopping for a brief moment before moving away are the key features of this category. Shadow appearance in the shadow category affects the performance and the foregrounds are not detected completely.

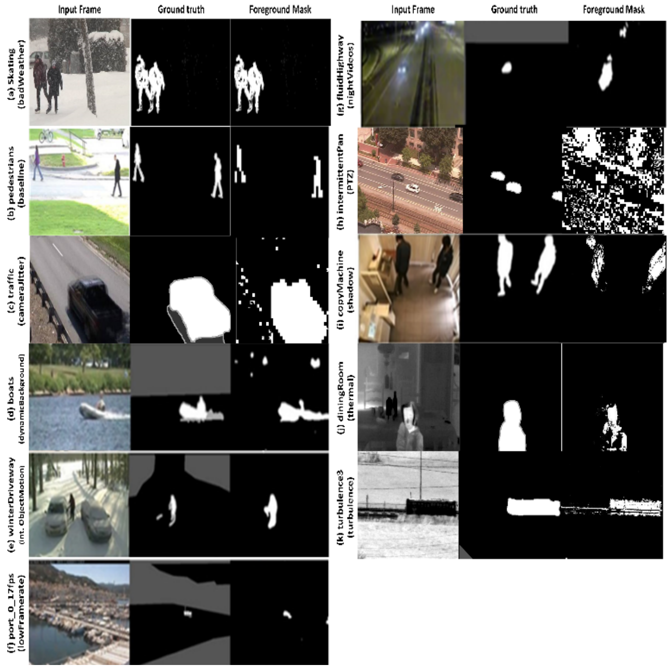

Input Frame
Ground truth
Foreground Mask
(a) Skating (badWeather)
(b) pedestrians (baseline)
(c) traffic (cameraJitter)
(d) boats (dynamicBackground)
(e) winterDriveway (int. ObjectMotion)
(f) port_0_17fps (lowFramerate)
Input Frame
Ground truth
Foreground Mask
(g) fluidHighway (nightVideos)
(h) intermittentPan (PTZ)
(i) copyMachine (shadow)
(j) diningRoom (thermal)
(k) turbulence3 (turbulence)

Fig. 4. Foreground Results of CDnet 2014 Dataset [39].

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

When we compared our results to different existing methods published on the CDnet website [39], we identified MOG [9], KNN [40], ViBe [12], and SubS [41] as candidates. Thus, we compared our proposed compressed-based background subtraction SDMBS with recent and state-of-the-art methods [26,42], classical methods like [9,40], and fast methods like the ViBe [12] Background Subtraction Algorithm.

In [26], a block-based MOG is designed to be processed using the compressed sensing CS elements of the frame-blocks CS-MOG and is targeted at WVSN applications, whereas [42] is a background model update algorithm that uses an intermittent technique along with an adaptive block-learning algorithm.

The results of three video sequences, Highway (baseline), Fountain2 (dynamic background), and Snowfall (bad weather), are illustrated in Fig. 5. The original video frame for the three datasets and its corresponding groundtruth are shown in the top two rows. The results of MOG [9], Vibe [12], two current state-of-the-art techniques [26, 42], and SDMBS are shown in the next five rows (from top to bottom). In the last row of Fig. 5, we demonstrate a qualitative comparison of our proposed technique with other current methods, revealing that our method outperforms several of the existing systems. From the results, it is observed that our system accurately recognises foreground objects and has a considerably high resemblance to the ground truth when compared to other examined systems.

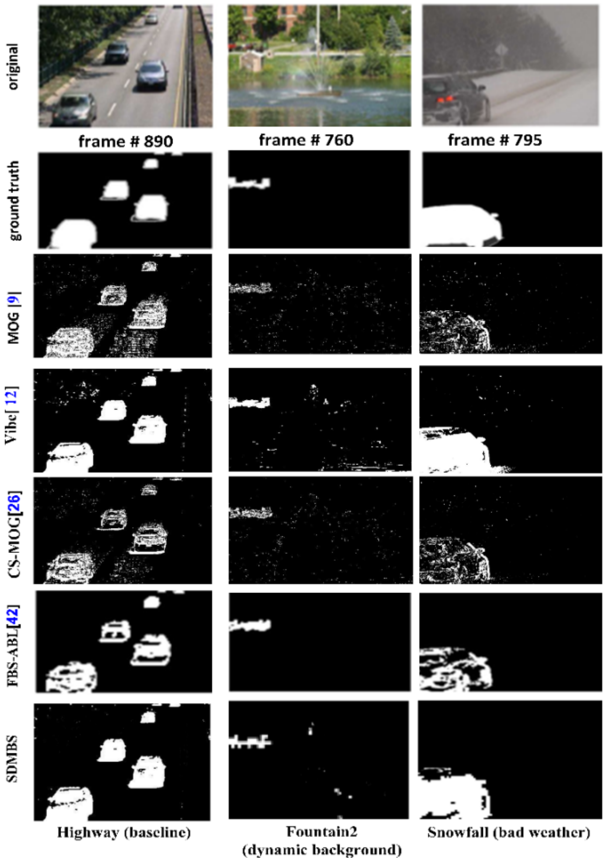


Fig. 5. Results with Highway (Baseline), Fountain2 (Dynamic Background), and Snowfall (Bad Weather) Videos Frames.

C. Quantitative Analysis

In the quantitative evaluation, our method is compared to widely popular and state-of-the-art object detection algorithms for WVSN by conducting experimentation on the benchmark (CDnet 2014) dataset [39]. Several evaluation metrics are utilised to provide a credible measure of the outcome. Average recall (Re), precision (Pr), and F-measure (Fm) for all the video sequences in each category are listed in Table I. True positive (TP), false positive (FP), true negative (TN), and false-negative (FN) are the four types of pixel-based count metrics that can be created using the available ground truth data [39].

As the frequency of false negatives decreases, the value of Recall (Re) increases, which is used to measure the degree of completeness of the recognised foreground. Precision (Pr) is a metric measuring how accurate the identified foreground is, with a lower value when there are a lot of false positives. F-measure (Fm) is a metric for determining the balance of recall and precision with equal weights, implying that it is high only when both recall and precision are high. The three evaluation metrics, recall (Re), precision (Pr), and F-measure (Fm), are only considered to avoid redundancy.

TABLE I. EVALUATION METRICS

Metrics	Description
Recall (Re)	$\frac{TP}{TP+FN}$
Precision (Pr)	$\frac{TP}{TP+FP}$
F-Measure (Fm)	$\frac{2(Pr.Re)}{Pr+Re}$
Specificity (SP)	$\frac{TN}{TN+FP}$
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$
False Negative Rate (FNR)	$\frac{FN}{TN+FP}$

The best and second-best performing approaches for each category, based on the average Fm for all the video sequences, are noted in red and bold in Tables II and III. When compared to classical methods, SDMBS may only be competitive in some areas, such as dynamic background, low frame rate, and bad weather. While there are approximate results for most categories with SubS [41] when SDMBS is ranked (2nd), this can be explained in terms of the design trade-off. While; when compared to state-of-the-art methods [26], we achieve a 15% increase in accuracy than CS-MOG [26] which is a compressed-based background subtraction applied for WVSN. In Fig. 7, the execution speed of SDMBS is compared to that of other methods at two resolution scales (320240 and 640480). For the two resolution scales, SDMBS excels in terms of speed. As seen in Fig. 7, SDMBS provides equivalent results to FBS-ABL [42], although it is more accurate, as seen in Fig. 6. When compared to other block-based techniques, this demonstrates SDMBS's effective design strategy.

TABLE II. COMPARISON ON THE FIRST SIX CATEGORIES OF CDNET 2014 DATASET.

Category	**Metrics**	**CDnet-14**	**MOG[9]**	**KNN[40]**	**ViBe[12]**	**SubS[41]**	**CS-MOG [26]**	**FBS-ABL [42]**	**SDMBS**
Baseline	**Re**	**0.9507**	0.8180	0.7934	0.8204	0.9520	0.7557	0.8910	0.8775
	Pr	**0.9347**	0.8461	0.9245	0.9288	0.9495	0.7942	0.8602	0.9481
	Fm	**0.9330**	0.8245	0.8411	0.8700	**0.9503**	0.7745	0.8649	**0.9114**
Dynamic background	**Re**	**0.8543**	0.8344	0.8047	0.7222	0.7768	0.6534	0.7958	0.7359
	Pr	**0.8606**	0.5989	0.6931	0.5346	0.8915	0.5262	0.7332	0.9604
	Fm	**0.8176**	0.6330	0.6865	0.5652	**0.8177**	0.583	0.7424	**0.8333**
Camera jitter	**Re**	**0.8159**	0.7334	0.7351	0.7375	0.8243	0.6826	0.8046	0.3281
	Pr	**0.8359**	0.5126	0.7018	0.4862	0.8115	0.4562	0.4656	0.5371
	Fm	**0.7806**	0.5969	**0.6894**	0.5720	**0.8152**	0.5469	0.5298	0.4074
Intermittent Object motion	**Re**	**0.7231**	0.5142	0.4617	0.5122	0.6578	0.4102	0.7861	0.5256
	Pr	**0.7888**	0.6688	0.7121	0.6515	0.7957	0.6012	0.7943	0.7639
	Fm	**0.6795**	0.5207	0.5026	0.5074	**0.6569**	0.48766	**0.7232**	0.6227
Shadow	**Re**	**0.9222**	0.7960	0.7478	0.7833	0.9419	0.7462	0.9143	ND
	Pr	**0.8551**	0.7156	0.7788	0.8342	0.8646	0.6366	0.8569	ND
	Fm	**0.8778**	0.7370	0.7468	0.8032	**0.8986**	0.687	**0.8671**	ND
Thermal	**Re**	**0.7727**	0.5691	0.4817	0.5435	0.8161	0.5131	0.6394	0.7277
	Pr	**0.8795**	0.8652	0.9186	0.9363	0.8328	0.8022	0.8002	0.8116
	Fm	**0.7962**	0.6621	0.6046	0.6647	**0.8171**	0.6258	0.6619	**0.7673**

TABLE III. COMPARISON ON THE NEWER CATEGORIES OF CDNET 2014 DATASET

Category	**Metrics**	**CDnet-14**	**MOG[9]**	**KNN[40]**	**ViBe[12]**	**SubS[41]**	**CS-MOG [26]**	**FBS-ABL [42]**	**SDMBS**
Low frame rate	**Re**	**0.7732**	0.5823	0.6290		0.8537	0.5323	0.6616	0.5934
	Pr	**0.6894**	0.6894	0.6865		0.6035	0.6394	0.7313	0.7398
	Fm	**0.6437**	0.5373	0.5491		**0.6445**	0.5809	0.6328	**0.6585**
Bad weather	**Re**	**0.7531**	0.7181	0.6537		0.8213	0.6881	0.7449	0.7978
	Pr	**0.8960**	0.7704	0.9114		0.9091	0.7354	0.8965	0.9385
	Fm	**0.8124**	0.7380	0.7587		**0.8619**	0.7109	0.8106	**0.8624**
Night videos	**Re**	**0.6107**	0.5261	0.5413		0.6570	0.4761	0.7498	0.6892
	Pr	**0.5438**	0.4128	0.4298		0.5359	0.3628	0.4957	0.4425
	Fm	**0.5154**	0.4097	0.4200		**0.5599**	0.4117	0.5272	**0.5386**
PTZ	**Re**	**0.7932**	0.6475	0.6980		0.8306	0.5975	0.8357	ND
	Pr	**0.3325**	0.1185	0.1979		0.2840	0.1685	0.2290	ND
	Fm	**0.3844**	0.1522	0.2126		**0.3476**	0.2628	**0.3267**	ND
Turbulence	**Re**	**0.7391**	0.7913	0.7682		0.8050	0.7413	0.9468	0.8023
	Pr	**0.7790**	0.4293	0.5117		0.7814	0.3793	0.4936	0.5392
	Fm	**0.7145**	0.4663	0.5198		**0.7792**	0.5018	0.5564	**0.6448**
CDnet 2014 average	**Re**	**0.7805**	0.6845	0.6649	0.6865	0.8124	0.6178	0.7972	0.6752
	Pr	**0.7543**	0.6025	0.6787	0.7286	0.7508	0.5547	0.6687	0.7423
	Fm	**0.7288**	0.5707	0.5937	0.6637	**0.7408**	0.5612	0.6584	**0.6940**

Fig. 6 and 7 highlight the trade-off between detection performance and execution speed, and as can be seen, extensively adaptable approaches have fast/practical execution at the cost of diminished performance. We achieve a 2.3x improvement in frame rate (FPS) over CS-MOG [26], a compressed-based background subtraction method used for WVSN. This shows an efficient decrease in processing time.

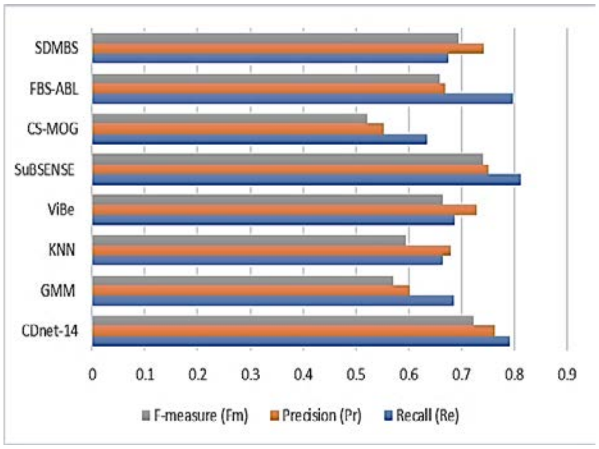


Fig. 6. Quantitative Analysis on CDnet Dataset.

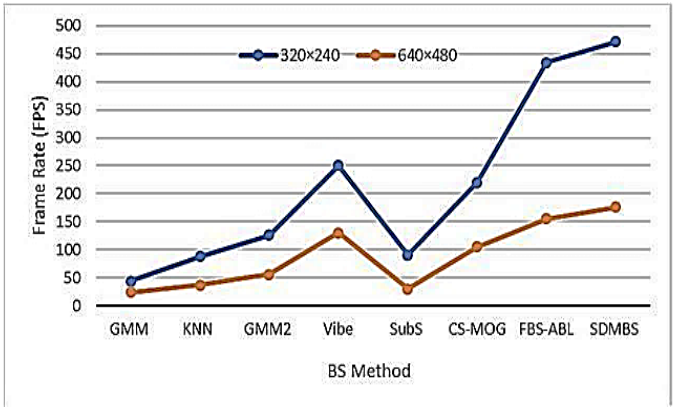


Fig. 7. Frame Rate for Different Techniques.

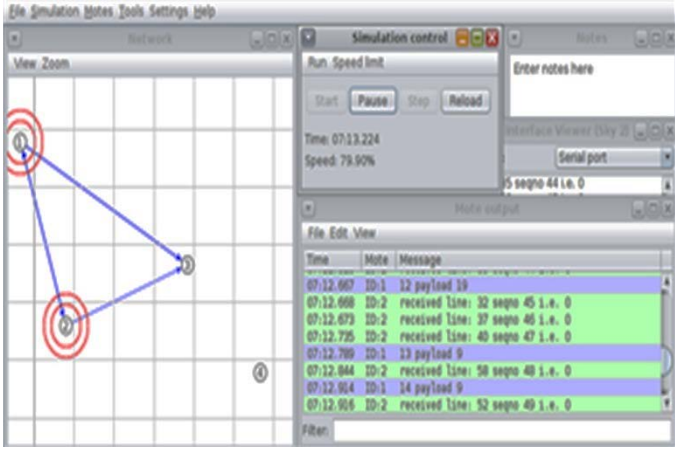
D. Sensor Network Simulation

This section illustrates the capability of the proposed system to work in WSN environments: first, simulation is carried out over Cooja of the Contiki OS Network Simulator [36], [37] to add the effect of lost packets and throughput. Second; the system is released on a real testbed using IOT-LAB [38]. Traffic trace files are used in the real testbed and simulated environment [15].

1) Cooja simulation: Four sensor nodes are installed. The sink is located at the left upper node (node 1) of the network area of 100 m × 100 m square grid. The destination node is located at (node 4). The simulation uses two datasets: pedestrians and PETS2006 (baseline) videos. The detection of moving objects is carried out at the host to select the blocks containing moving objects, and the blocks are then sent and routed through intermediate nodes to the base station. The received blocks at the destination are reconstructed to show the moving target, Fig. 8.

Fig. 9 shows the received image PSNR for two approaches: First, the full transmission of the image frame (Full Tx), while the second is our approach to transmitting the important portion of the image containing the moving object (Partial Tx). Although the proposed approach has a lower PSNR ratio than the full transmission approach, however, the average value is 27db. PSNR and energy are calculated using [15].

The proposed approach was compared to the direct approach for the energy consumption analysis. In the direct technique, a multi-hop transfer of an entire image to the sink node is used. On the aforementioned datasets, Fig. 10 depicts the energy using both methodologies. As can be observed, the node's energy usage has been significantly reduced. The two datasets show that the suggested approach can be employed in real-time moving object detection systems since a portion of the image data, including object information, is received at the sink node with an appropriate range of PSNR values and less energy.



Enter notes here

Time: 07:13.224
Speed: 79.90%

5 segno 44 i.e. 0

```
Time       Mote  Message
07:12.667  ID:1  12 payload 19
07:12.668  ID:2  received line: 32 segno 45 i.e. 0
07:12.673  ID:2  received line: 37 segno 46 i.e. 0
07:12.735  ID:2  received line: 40 segno 47 i.e. 0
07:12.789  ID:1  13 payload 9
07:12.844  ID:2  received line: 58 segno 48 i.e. 0
07:12.914  ID:1  14 payload 9
07:12.916  ID:2  received line: 52 segno 49 i.e. 0
```

Fig. 8. Cooja Snapshot.

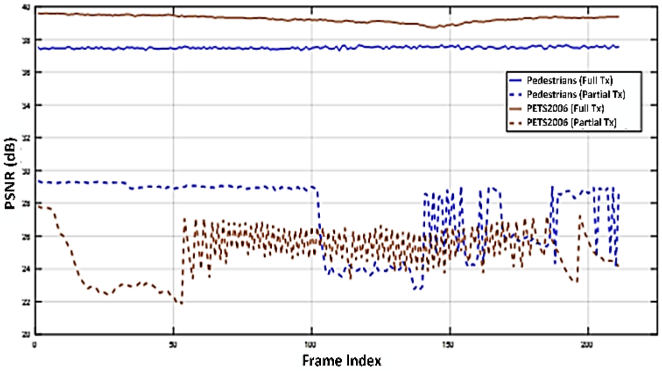


Fig. 9. PSNR for the Two Datasets.

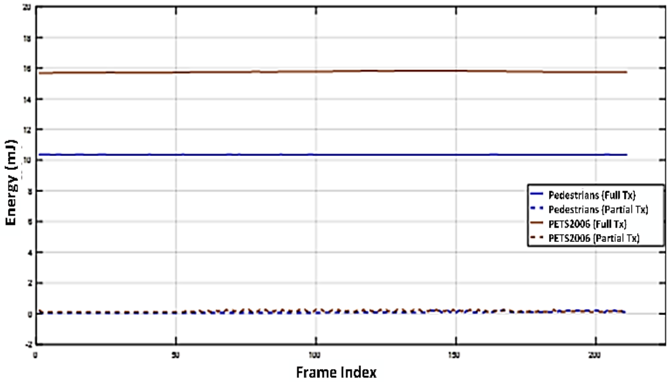


Fig. 10. Energy Consumption.

2) IOT-LAB realization: IoT-LAB[2] [38] is a large-scale WSN testbed that includes over 2000 wireless sensor nodes and a variety of processor architectures and wireless chips. IoT-LAB can be accessed through a web portal or by using the command-line tools. It allows users to retrieve experiment results and access serial ports on devices. Based on trace files as presented in [15], the IoT-LAB testbed M3 open nodes illustrated in Fig. 11(b) was employed in our experiments to replicate the intended object detection of the two datasets: pedestrians and PETS2006 (baseline). As shown in Fig. 11(a), the nodes m3-1, m3-10, m3-15, and m3-16 are used as senders, and m3-24 (blue circles) is used as a receiver to acquire varied loss rates as shown in Fig. 11(a). The sender (sender node) sends data packets according to the sender's trace file specifications (st-packet). The receiver (receiver node) maintains track of the packets it receives in a receiver trace file (rt-packet) as shown in Fig. 11. The sequence numbers of correctly received packets are received on the user's computer, which is used to reconstruct the video and calculate experiment metrics.

Fig. 12 shows the results of applying the proposed moving object detection technique in IOT-LAB to the two datasets: pedestrians on the first row and PETS2006 on the second row. The foreground blocks are transmitted and routed to the sink node. The sink node decompresses the received block and determines the moving object's location. For surveillance applications, object location is the most important piece of information that requires further analysis. The object ROI is transmitted to the sink node correctly with minimum network resources, memory, and bandwidth. The energy is minimised with an accepted PSNR.

V. Conclusion

A background subtraction method that is both computationally efficient and accurate has been developed for object tracking across the limited resources of Wireless Visual Sensor Networks (WVSNs). To address the computation bottleneck of processing for constrained sensor networks, we use partial DCT to reduce the data dimensions while preserving the information content. In addition, energy-efficient block-based dual-mode SGM is utilised for foreground block detection, where the image frame is divided into blocks and only blocks containing foreground pixels are further processed for the refining stage. The foreground pixels are determined and the moving object is located. In contrast to standard Compress Sensing CS, which compresses the entire frame, the target region of interest ROI in our proposed method is compressed, communicated, and routed toward the sink node for further analysis. Our experimental results show that our method is as efficient as traditional algorithms. Moreover, it is up to three times faster than the state-of-the-art WVSN object detection methods, and 15% more accurate. For embedded camera networks, we demonstrate that our suggested technique can accurately detect a moving object in real-time. We applied the proposed detection method in a WSN environment using Cooja of the Contiki OS Network Simulator. We verified that the energy required for transmitting the detected object to the sink node in our proposed detection method is lower than that of comparable methods at acceptable PSNRs. Finally; the system is released on a real testbed using IOT-LAB using testbed M3 open nodes.

[2] https://www.iot-lab.info/

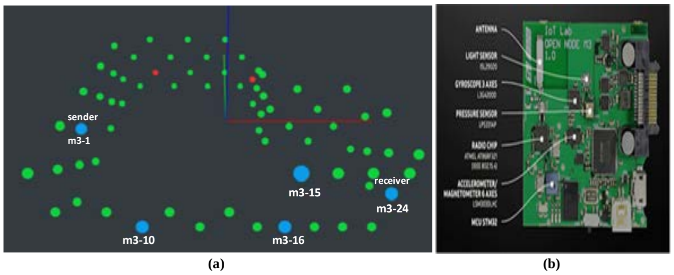


Fig. 11. IoT-LAB (a) Configuration, (b) Testbed M3 Open Nodes (ARM Cortex M3, 32-Bits MCU, and 802.15.4 PHY Standard).

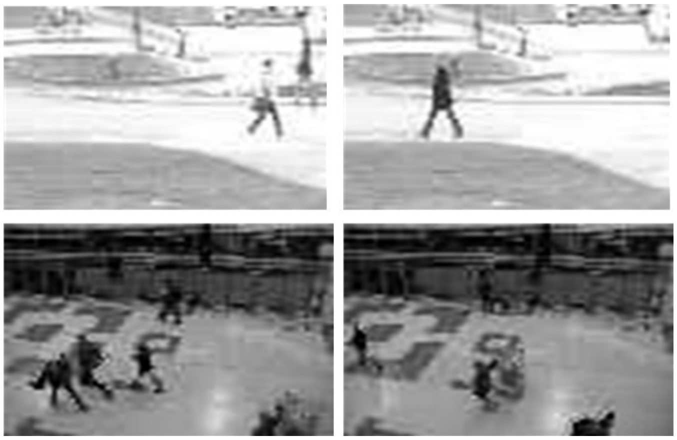


Fig. 12. Object Detection Received at Destination Node Sink for the Two Datasets: Pedestrians (Upper Row) and PETS2006 (Lower Row).

References

[1] T.C. H K Patil, Wireless Sensor Networks - an overview | ScienceDirect Topics, Am. Sci. Res. J. Eng. Technol. Sci. 64 (2017).https://www.sciencedirect.com/topics/computer-science/wireless-sensor-networks.

[2] I.F. Akyildiz, T. Melodia, K.R. Chowdury, Wireless multimedia sensor networks: A survey, *IEEE Wirel. Commun*. 14 (2007) 32–39. https://doi.org/10.1109/MWC.2007.4407225.

[3] Y. Ye, S. Ci, A.K. Katsaggelos, Y. Liu, Y. Qian, Wireless video surveillance: A survey, *IEEE Access*. 1 (2013) 646–660. https://doi.org/10.1109/ACCESS.2013.2282613.

[4] B. Ma, L. Huang, J. Shen, L. Shao, M.H. Yang, F. Porikli, Visual Tracking under Motion Blur, *IEEE Trans. Image Process*. 25 (2016) 5867–5876. https://doi.org/10.1109/TIP.2016.2615812.

[5] B. Garcia-Garcia, T. Bouwmans, A.J.R. Silva, Background subtraction in real applications: Challenges, current models and future directions, *Comput. Sci. Rev*. 35 (2020). https://doi.org/10.1016/j.cosrev.2019.100204.

[6] X. Li, M.K. Ng, X. Yuan, Median filtering-based methods for static background extraction from surveillance video, Numer. Linear Algebr. with Appl. 22 (2015) 845–865. https://doi.org/10.1002/nla.1981.

[7] S. Maity, A. Chakrabarti, D. Bhattacharjee, Block-Based Quantized Histogram (BBQH) for efficient background modeling and foreground extraction in video, in: 2017 *Int. Conf. Data Manag. Anal. Innov. ICDMAI* 2017, 2017: pp. 224–229. https://doi.org/10.1109/ICDMAI.2017.8073514.

[8] M. Boninsegna, A. Bozzoli, Tunable algorithm to update a reference image, Signal Process. Image Commun. 16 (2000) 353–

365. https://doi.org/10.1016/S0923-5965(99)00063-6.

[9] C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2 (1999) 246–252. https://doi.org/10.1109/cvpr.1999.784637.

[10] Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2000: pp. 751–767. https://doi.org/10.1007/3-540-45053-x_48.

[11] K. Kim, T.H. Chalidabhongse, D. Harwood, L. Davis, Real-time foreground-background segmentation using codebook model, *Real-Time Imaging*. 11 (2005) 172–185. https://doi.org/10.1016/j.rti.2004.12.004.

[12] O. Barnich, M. Van Droogenbroeck, ViBe: A universal background subtraction algorithm for video sequences, *IEEE Trans. Image Process*. 20 (2011) 1709–1724. https://doi.org/10.1109/TIP.2010.2101613.

[13] D.L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory*. 52 (2006) 1289–1306. https://doi.org/10.1109/TIT.2006.871582.

[14] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. 23, pp. 90–93, 1974.

[15] M. Maimour, SenseVid: A traffic trace based tool for QoE Video transmission assessment dedicated to Wireless Video Sensor Networks, *Simul. Model. Pract. Theory*. 87 (2018) 120–137. https://doi.org/10.1016/j.simpat.2018.06.006.

[16] R. Banerjee, S. Das Bit, Low-overhead video compression combining partial discrete cosine transform and compressed sensing in WMSNs, *Wirel. Networks*. 25 (2019) 5113–5135. https://doi.org/10.1007/s11276-019-02119-y.

[17] K.M. Yi, K. Yun, S.W. Kim, H.J. Chang, H. Jeong, J.Y. Choi, Detection of moving objects with non-stationary cameras in 5.8ms: Bringing motion detection to your mobile device, in: *IEEE Comput*. Soc. Conf. Comput. Vis. Pattern Recognit. Work., 2013: pp. 27–34. https://doi.org/10.1109/CVPRW.2013.9.

[18] F. De La Torre, M.J. Black, A framework for robust subspace learning, Int. J. Comput. Vis. 54 (2003) 117–142. https://doi.org/10.1023/A:1023709501986.

[19] Xiaowei Zhou, Can Yang, Weichuan Yu "Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35:597-610, 2013.

[20] A. Zheng, T. Zou, Y. Zhao, B. Jiang, J. Tang, C. Li, Background subtraction with multi-scale structured low-rank and sparse factorization, *Neurocomputing*. 328 (2019) 113–121. https://doi.org/10.1016/j.neucom.2018.02.101.

[21] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory*. 53 (2007) 4655–4666. https://doi.org/10.1109/TIT.2007.909108.

[22] M. Lamarre, J.J. Clark, Background subtraction using competing models in the block-DCT domain, in Proc. - *Int. Conf. Pattern Recognit*., 2002: pp. 299–302.https://doi.org/10.1109/ICPR.2002.1044695.

[23] Y.A. Ur Rehman, M. Tariq, T. Sato, A novel energy efficient object detection and image transmission approach for wireless multimedia sensor networks, *IEEE Sens. J*. 16 (2016). https://doi.org/10.1109/JSEN.2016.2574989.

[24] D.M. Pham, S.M. Aziz, Object extraction scheme and protocol for energy efficient image communication over wireless sensor networks, *Comput. Networks*. 57 (2013) 2949–2960. https://doi.org/10.1016/j.comnet.2013.07.001.

[25] S.A. Nandhini, S. Radha, R. Kishore, Efficient compressed sensing based object detection system for video surveillance application in WMSN, *Multimed. Tools Appl*. 77 (2018) 1905–1925. https://doi.org/10.1007/s11042-017-4345-2.

[26] Y. Shen, W. Hu, M. Yang, J. Liu, B. Wei, S. Lucey, C.T. Chou, Real-time and robust compressive background subtraction for embedded camera networks, *IEEE Trans. Mob. Comput*. 15 (2016) 406–418. https://doi.org/10.1109/TMC.2015.2418775.

[27] A. Tulsyan, B. Huang, R.B. Gopaluni, J.F. Forbes, Performance assessment, diagnosis, and optimal selection of non-linear state filters, *J. Process Control*. 24 (2014) 460–478. https://doi.org/10.1016/j.jprocont.2013.10.015.

[28] M. Khare, R.K. Srivastava, A. Khare, Moving object segmentation in Daubechies complex wavelet domain, *Signal, Image Video Process*. 9 (2015) 635–650. https://doi.org/10.1007/s11760-013-0496-4.

[29] S.S. Sengar, S. Mukhopadhyay, Moving object detection using statistical background subtraction in wavelet compressed domain, *Multimed. Tools Appl*. 79 (2020) 5919–5940. https://doi.org/10.1007/s11042-019-08506-z.

[30] W. Wang, J. Yang, W. Gao, Modeling background and segmenting moving objects from compressed video, *IEEE Trans. Circuits Syst. Video Technol*. 18 (2008) 670–681. https://doi.org/10.1109/TCSVT.2008.918800.

[31] S. Popa, D. Crookes, P. Miller, Hardware acceleration of background modeling in the compressed domain, *IEEE Trans. Inf. Forensics Secur*. 8 (2013) 1562–1574. https://doi.org/10.1109/TIFS.2013.2276753.

[32] H. Ye, J. Pei, Infrared images target detection based on background modeling in the discrete cosine domain, in: 2018: p. 33. https://doi.org/10.1117/12.2285785.

[33] P. Zhang, X. Wang, X. Wang, C. Fei, Z. Guo, Infrared small target detection based on spatial-temporal enhancement using quaternion discrete cosine transform, *IEEE Access*. 7 (2019) 54712–54723. https://doi.org/10.1109/ACCESS.2019.2912976.

[34] S. Manimozhi, S. Aasha Nandhini, S. Radha, Compressed Sensing based background subtraction for object detection in WSN, in: 2015 Int. Conf. Commun. Signal Process. ICCSP 2015, 2015: pp. 569–573. https://doi.org/10.1109/ICCSP.2015.7322550.

[35] S.W. Kim, K. Yun, K.M. Yi, S.J. Kim, J.Y. Choi, Detection of moving objects with a moving camera using non-panoramic background model, Mach. Vis. Appl. 24 (2013) 1015–1028. https://doi.org/10.1007/s00138-012-0448-y.

[36] A. Dunkels, B. Grönvall, T. Voigt, Contiki - A lightweight and flexible operating system for tiny networked sensors, in: Proc. - *Conf. Local Comput. Networks*, LCN, 2004: pp. 455–462. https://doi.org/10.1109/LCN.2004.38.

[37] F. Österlind, A. Dunkels, J. Eriksson, N. Finne, T. Voigt, Cross-level sensor network simulation with COOJA, in: Proc. - *Conf. Local Comput. Networks*, LCN, 2006: pp. 641–648. https://doi.org/10.1109/LCN.2006.322172.

[38] C. Adjih, E. Baccelli, E. Fleury, G. Harter, N. Mitton, T. Noel, R. Pissard-Gibollet, F. Saint-Marcel, G. Schreiner, J. Vandaele, T. Watteyne, FIT IoT-LAB: A large scale open experimental IoT testbed, in: *IEEE World Forum Internet Things*, WF-IoT 2015 - Proc., 2015: pp. 459–464. https://doi.org/10.1109/WF-IoT.2015.7389098.

[39] Y. Wang, P.M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, P. Ishwar, CDnet 2014: An expanded change detection benchmark dataset, in: *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. Work., 2014: pp. 393–400. https://doi.org/10.1109/CVPRW.2014.126.

[40] Z. Zivkovic, F. Van Der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction, *Pattern Recognit. Lett*. 27 (2006) 773–780. https://doi.org/10.1016/j.patrec.2005.11.005.

[41] P.L. St-Charles, G.A. Bilodeau, R. Bergevin, SuBSENSE: A universal change detection method with local adaptive sensitivity, *IEEE Trans. Image Process*. 24 (2015) 359–373. https://doi.org/10.1109/TIP.2014.2378053.

[42] V.J. Montero, W.Y. Jung, Y.J. Jeong, Fast background subtraction with adaptive block learning using expectation value suitable for real-time moving object detection, *J. Real-Time Image Process*. 18 (2021) 967–981. https://doi.org/10.1007/s11554-020-01058-8.