



**HAL**  
open science

## An efficient low complexity region-of-interest detection for video coding in wireless visual surveillance

Ahcen Aliouat, Nasreddine Kouadria, Saliha Harize, Moufida Maimour

### ► To cite this version:

Ahcen Aliouat, Nasreddine Kouadria, Saliha Harize, Moufida Maimour. An efficient low complexity region-of-interest detection for video coding in wireless visual surveillance. *IEEE Access*, 2023, 11, pp.26793-26806. 10.1109/ACCESS.2023.3248067 . hal-04044211

**HAL Id: hal-04044211**

**<https://hal.science/hal-04044211>**

Submitted on 25 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

# An Efficient Low Complexity Region-of-Interest Detection for Video Coding in Wireless Visual Surveillance

AHCEN ALIOUAT, (Student Member, IEEE) <sup>1</sup>, NASREDDINE KOUADRIA <sup>1</sup>, SALIHA HARIZE <sup>1</sup>, AND MOUFIDA MAIMOUR <sup>2</sup>

<sup>1</sup>LASA Laboratory, Badji Mokhtar University of Annaba, Faculty of Technology, Department of Electronics, Annaba 23000, Algeria

<sup>2</sup>Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

Corresponding author: Ahcen Aliouat (e-mail: ahcen.aliouat@ieee.org).

This work was partially supported by Campus France 46082TB and PHC TASSILI program 21MDU323.

**ABSTRACT** Following its use in several applications, including video coding in wireless surveillance, moving object detection (MOD) has become a popular video analysis topic. Despite the considerable progress in the accuracy of MOD for video coding, its implementation in constrained sensors is a real challenge owing to their high complexity and energy consumption. Therefore, there is a great need to address the trade-off between the accuracy and the energy efficiency of MOD approaches for video coding in constrained systems. In this work, an energy-efficient region-of-interest (ROI) detection algorithm as a pre-encoder for wireless visual surveillance (WVS) is proposed. The algorithm ensures a trade-off between detection accuracy and computational complexity. To this end, we propose constructing an activity map by measuring each block activity between successive frames. The map scores are processed using a combination of a fast Gaussian smoother and a rank-order filter to improve accuracy. Only the blocks in motion are coded and counted for transmission. The accuracy of our approach has been evaluated on a large dataset using key performance metrics. It has been found that our algorithm outperforms other state-of-the-art techniques in terms of true positive rate (TPR), with 80.84% on sensitivity metric, while exhibiting a well-balanced accuracy for all categories. A careful examination of the computational complexity confirms the low overhead. The energy and bitrate savings could achieve nearly 90% and 98%, respectively.

**INDEX TERMS** region-of-interest, object detection, image compression, WVS, video surveillance, energy-efficiency

## I. INTRODUCTION

VIDEO Content Analysis (VCA) techniques involve automatically analyzing video to detect and determine spatial and temporal events. VCA is used in a wide range of domains, including video browsing and retrieval [1], image and video coding [2] [3], video surveillance, etc. The analyze-then-compress (ATC) paradigm employs VCA to first analyze the content before compressing it. As a result, feature extraction is carried out before the compression and the transmission of the visual data captured from a visual sensor node (VS). This paradigm has been put forth as a substitute for the conventional compress-then-analyze (CTA) paradigm, which compresses the entire captured video before transmitting it to be processed further when it is received. The CTA paradigm typically employs highly complex video

coding standards [4], such as MJPEG [5], H264 [6], and HEVC [7]. As a result, the ATC can streamline this process and enable video coding within limited resources devices [2].

Because only a few parts of the frame, called region-of-interest (ROI), are compressed and transmitted, the ATC paradigm [8] is suitable for many applications [9]–[11]. In such applications, the end-user is only interested in the ROI, and therefore it would be relevant to extract those regions before encoding. This enables the development and employment of very low-bitrate encoders. Additionally, the energy the sensor node uses during transmission is frequently larger than the energy used for processing, namely, compression [12]. Hence, transmitting only the ROI by sacrificing some image quality reduces the bitstream, which allows for a high amount of bandwidth and energy savings. The framework of

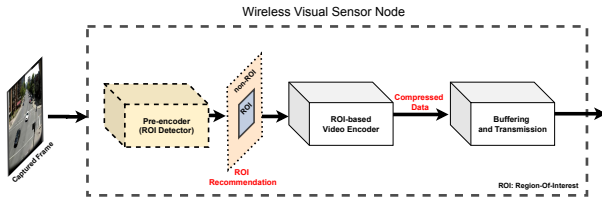


FIGURE 1: Image analysis by ROI detection for video coding in WVS

this paradigm is shown in Figure (1).

Despite the increased advancement in new video coding standards involving high video quality with very low bitrates, they are still not adopted for WVS [13] [14]. The unsuitability is due to the high complexity of the used coding modules, either in the intra or inter-coding modes. New approaches and paradigms have emerged to overcome this problem. They aim to code the frames based on the ROI and the difference between successive frames [15] [16]. Correspondingly, the moving object is the salient zone in the frame that must be coded, whereas non-ROI could be omitted to save bandwidth and energy in an ROI-based energy optimization approach. ROI is an important element in this context; therefore, accurate ROI detection is a crucial step that must be well-studied.

Making the tradeoff between accurate object detection and very low complexity is an important subject to be addressed and studied to advance the ROI-based video coding paradigm. To minimize contextual loss, this tradeoff must address the detection of the entire ROI (high sensitivity), but with a moderate energy budget [17] [18] [19]. Indeed, the benefit of using those approaches in video surveillance is that the cameras deployed are relatively stable. As a result, there are a few insignificant background changes. Accordingly, this topic has received a lot of attention in recent years. There have been numerous pre-encoder approaches for video and image compression in WVS [20] [21] [22] [23].

In this work, we address the proposition of an energy-efficient pre-encoder for WVS. The contributors address the combination and evaluation of simple but efficient techniques that have not been addressed previously within the scope of object-based video coding for WVS. We contribute by proposing an energy-efficient method and validating its detection efficiency on a large dataset containing nearly all surveillance conditions classified into 11 categories [24]. Additionally, we validate the energy efficiency through detailed modeling of computational complexity and energy cost to prove the neglected extra cost compared to the saved energy. By avoiding unnecessarily processed and compressed blocks, the proposed pre-encoding scheme significantly reduces computational complexity.

A Block-based movIng Region Detection (BIRD) technique is proposed. BIRD detects the difference created between frames using a kind of Sum of absolute Frames Difference (SFD) [11] to address video coding in resource-constrained systems. The SFD operation is followed by

heavy yet efficient morphological filtering to enhance the accuracy of the moving-region detection. A threshold is used to extract the binary mask of the moving region. The framework is considered an efficient first step in an ATC paradigm. Contrary to compressing and transmitting the whole frame (i.e CTA), the proposed approach enables compression and transmission of only the activity blocks. This method will drastically decrease the processing and transmission energy budget in a WVS while maintaining an acceptable quality of service (QoS) and a high frame rate. The main contributions of the proposal are as follows:

- A low complexity ROI detection method dedicated to video coding in constrained wireless surveillance systems.
- The detection accuracy is improved through a combination of a fast Gaussian smoother and a rank-order filter.
- The algorithm is assessed using several metrics to evaluate the detection performance and confirm its superiority compared to the state-of-the-art techniques in constrained wireless surveillance systems.
- Bitrate and energy savings are achieved using the algorithm as a pre-encoder of a baseline JPEG compression chain.
- Based on an energy/memory consumption modeling, using ARM Cortex M3 characteristics, the viability of the algorithm is demonstrated for implementation in WVS.

The remainder of this paper is organized as follows. A background and related work review is presented in Section II. The proposed algorithm is presented in detail in Section III. Section IV shows the results and evaluation of the proposed method in terms of detection accuracy, complexity, energy, speed, and memory performance. Finally, a conclusion is drawn in Section V.

## II. BACKGROUND AND RELATED WORK

The literature has extensively discussed and analyzed the design of energy-efficient Wireless Sensor Networks (WSNs) [25] [26] [27] [28]. The approaches vary depending on whether the contributions are in the processing, the transmission, or the network part. The recommended solutions often focus on identifying resource allocation techniques that use the least amount of energy. The resource under consideration can comprise memory usage, data compression algorithms, data routing, and transmission power at the radio part. A kind of WSN that integrates a camera sensor is known as a Wireless Multimedia Sensor Network (WMSN), which incorporates extensive use of the resources. WMSN's resources are exhausted extensively due to the amount of multimedia data (i.e.: images and video). There is a real need to reduce the amount of captured data intelligently with a minimum loss to enhance the efficiency of WMSN. It is essential to make a tradeoff between the added energy cost of the data reduction technique, the final gain in energy from its implementation, and the QoS degradation. Many approaches have been proposed in this context to achieve the

TABLE 1: Summary of the related work on ROI-based video coding

Algorithm	Methodology	Highlights	Limitations
Kouadria et al. (2019) [22]	<ul style="list-style-type: none"> <li>• <math>8 \times 8</math> SAD</li> <li>• thresholding to extract ROI mask.</li> <li>• DTT transform for compression</li> </ul>	<ul style="list-style-type: none"> <li>• low complexity</li> <li>• fast image compression algorithm</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• less accurate</li> <li>• few datasets</li> <li>• few evaluation metrics</li> </ul>
Rehman et al. (2016) [29]	<ul style="list-style-type: none"> <li>• divide the frame into 4 blocks</li> <li>• select ROI from sub-blocks</li> <li>• background modeling</li> <li>• compression using DWT</li> </ul>	<ul style="list-style-type: none"> <li>• moderate accurate detection</li> <li>• simple and efficient algorithm</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• limited datasets</li> <li>• high bitrate</li> <li>• high complexity for WMSN node</li> </ul>
Aliouat et al. (2022) [30]	<ul style="list-style-type: none"> <li>• edge detection using Canny filter</li> <li>• <math>8 \times 8</math> SAD of the edge map</li> <li>• automatic multi-threshold selection</li> <li>• multi-Otsu thresholding</li> <li>• compression priority to the ROI</li> </ul>	<ul style="list-style-type: none"> <li>• automatic thresholding</li> <li>• accurate detection</li> <li>• content-aware coding</li> <li>• allocate more resources to the ROI</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• high complexity</li> <li>• limited dataset</li> <li>• high bitrate (50% reduction)</li> <li>• no energy consumption model</li> <li>• few evaluation metrics</li> </ul>
Aliouat et al. (2022) [16]	<ul style="list-style-type: none"> <li>• edge detection using Sobel filter</li> <li>• <math>4 \times 4</math> SAD of the edge map</li> <li>• 2-D Rank order map filtering</li> <li>• fixed threshold</li> <li>• background update each GOP</li> </ul>	<ul style="list-style-type: none"> <li>• good accuracy on the used dataset</li> <li>• efficient in different weather cond.</li> <li>• high bitrate and processing reduction</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• high complexity for WMSN context</li> <li>• limited dataset</li> <li>• no energy consumption model</li> <li>• few evaluation metrics</li> </ul>
Ko. et al. (2018) [23]	<ul style="list-style-type: none"> <li>• edge detection using Sobel filter</li> <li>• <math>8 \times 8</math> SAD</li> <li>• bitrate control using PID-controller</li> <li>• optimal enhancement algorithm</li> <li>• prototyping on 130nm sensor node.</li> <li>• FPGA implementation</li> </ul>	<ul style="list-style-type: none"> <li>• accurate detection</li> <li>• optimal circuit design</li> <li>• high processing and bitrate reduction</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• limited dataset (2 sequences)</li> <li>• no comparison to the state of the art</li> <li>• few evaluation metrics</li> </ul>
Ko. et al. (2015) [31]	<ul style="list-style-type: none"> <li>• edge detection using Sobel filter</li> <li>• perform Frame difference</li> <li>• <math>8 \times 8</math> SAD</li> <li>• rate control (channel cond. -BER-)</li> <li>• thresholding using PID controller</li> </ul>	<ul style="list-style-type: none"> <li>• optimal circuit design</li> <li>• high processing and bitrate reduction</li> <li>• content and energy-aware</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• limited dataset (4 sequences)</li> <li>• no comparison to the state-of-the-art</li> <li>• few evaluation metrics</li> <li>• detection accuracy not reported</li> </ul>
Aliouat et al. (2023) [32]	<ul style="list-style-type: none"> <li>• novel (S-SAD) introduced</li> <li>• multi-classes coding 2 based on ROI.</li> <li>• assessed for Human and Machine based monitoring</li> </ul>	<ul style="list-style-type: none"> <li>• accurate detection</li> <li>• energy model provided</li> <li>• high bitrate and processing saving</li> <li>• content-awareness</li> <li>• resources/quality tradeoff achieved</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• no detection accuracy comparison</li> <li>• medium dataset</li> <li>• fixed threshold</li> </ul>
Sengar et al. (2020) [33]	<ul style="list-style-type: none"> <li>• MOD detection using Optical flow</li> <li>• Ostu for thresholding</li> <li>• particle swarm optimization (PSO) for redundancy exploring</li> </ul>	<ul style="list-style-type: none"> <li>• deal with moving cameras</li> <li>• good efficiency comparison with state-of-the-art</li> <li>• good rate-distortion performance</li> </ul>	<ul style="list-style-type: none"> <li>• limited dataset (4 sequences)</li> <li>• no energy consumption model</li> <li>• not dedicated to WMSN context</li> <li>• few evaluation metrics</li> </ul>
BIRD (Proposed)	<ul style="list-style-type: none"> <li>• <math>8 \times 8</math> SFD</li> <li>• 1-D ROF on the activity map</li> <li>• FGS filter on the activity map</li> <li>• a pre-encoder for video coding.</li> </ul>	<ul style="list-style-type: none"> <li>• low complexity</li> <li>• high detection accuracy</li> <li>• energy modeling (ARM Cortex M3)</li> <li>• large dataset (51 sequences)</li> <li>• dedicated to WMSN context</li> </ul>	<ul style="list-style-type: none"> <li>• tested only for fixed camera</li> <li>• fixed threshold</li> </ul>

intended target by using low-cost and classical techniques or advanced techniques based on machine learning (ML) and deep learning (DL) [34].

One of the research axes is to use feature extraction as a data reduction technique in WMSN [35]. The authors in [36] use the FAST and the BRIEF algorithms for image feature extraction and matching. Likewise, the work in [8] proposes a method for visual-features compression for WMSN based on the ATC paradigm. Alongside the intensive complexity of extracting visual features, the content of the video could not be reconstructed in the pixel domain, which is still a major drawback of feature extraction-based video coding.

For visual data compression, other methods rely on movement detection. The moving object extraction in this scenario is a crucial step. For instance, by building the background model using background subtraction (BS), the moving object could be detected. The background models are obtained by applying well-known methods such as GMM [37], Histogram of Gradient (HoG) [38], codebook [39], and ViBe [40] or using deep learning-based techniques [41] [42]. However, DL-based techniques are not suitable for some special scenarios and systems, especially for non-powerful computing abilities. Despite their good performances in MOD tasks, the aforementioned techniques are energy expensive, making them unsuitable for embedded nodes.

The alternative to these techniques is to use simple yet efficient MOD techniques, such as frame difference (FD) and BS [43] [44]. FD has been used for MOD and has presented advantages because of its low complexity, low memory, and processing speed. However, it has low accuracy when dealing with noisy backgrounds [31]. Edge Detection (ED) has also provided solutions to enhance the efficiency of the MOD algorithms, but it could suffer from high computational costs due to the edge detection operator's calculations. As a result, An inexpensive ED operator is required [30] [31].

In [22], the authors proposed an ROI-based image coding technique. The idea is to detect, compress and send only the moving blocks in the frame while a low-cost compression technique is used by applying the integer discrete Tchebichef transform. In [29], another method has been proposed by Rehman et al., which involves dividing the frame into four main blocks and detecting the moving object on each block using a probabilistic approach. The transmission is subsequently limited to the moving segments after compression using the wavelet transform-based compression approach. In [33], the authors proposed a surveillance video compression based on motion detection and segmentation, employing a JPEG-like chain to compress the data.

In [16], the authors proposed an ED-based ROI detection technique using the Sobel edge detector to extract the edges of the moving regions and create an activity map based on those detected edges. While in [30], the Canny ED is used as a low-cost edge detection method to extract the ROI prior to compression. ROI detection has also been a solution to control the memory usage [45], the bitrate [46], and the quality of the video encoder. This is accomplished by managing the bit

allocation mechanisms, as shown in [47]. The authors in [32] have provided a more accurate and energy-efficient strategy, in which a good trade-off between energy efficiency, image quality, content awareness, bitrate, and effective machine-based monitoring tasks at the destination have been reached. The strategy seeks to create a new pre-processing method, named Successive Sum of Absolute Differences (S-SAD), to identify the ROI and divide it into many classes based on their importance. Table 1 summarizes the related work on ROI-based video coding for WVS.

While the works mentioned above provide effective energy-saving solutions for WVS systems, the majority of them did not fully consider the efficiency of the used moving object detection method because they only considered a few evaluation metrics and small datasets. The presented works did not address evidence of the effectiveness of the ROI detection techniques used. Furthermore, some methods are weakened by the high amount of data that must be transmitted, making them less efficient under the WMSN constraints [12]. The accuracy of the detection of the object is crucial. If a high level of detection accuracy is not guaranteed, complex distortions during frame reconstruction will appear.

The above-mentioned brief review reveals the fact that numerous researchers are devoted to investigating ROI-based video coding in WMSN. The techniques used produce varying degrees of accuracy and complexity. However, to the best of our knowledge, there is no literature validating a good accuracy-complexity tradeoff of the moving object detection techniques for ROI-based video compression in WVS. In this article, a tradeoff between accuracy and complexity is achieved by the proposed BIRD. The assumptions are validated through an application on a large dataset and an energy and memory consumption model.

### III. PROPOSED METHOD

The main purpose of the BIRD method is the exploitation of the successive changes between two frames  $F_n$  and  $F_m$ , with  $m < n$ , where  $n$  and  $m$  are respectively the current and a previous frame in the captured video. The frame difference method is of very low complexity and simple to implement, which makes it an appropriate choice to suit the constrained resources in a WSN. Meanwhile, it suffers from low region detection accuracy [31]. To overcome the low accuracy of pixel-based detection of the frame difference method, the blocks of the resulting difference are summed up to create an activity map that represents the level of the activity in each region.

#### A. DIFFERENCE DETECTION

Let  $\phi_n$  and  $\phi_m$  be the intensity map of the frames  $F_n$  and  $F_m$  of the size  $M \times N$ . Based on the SFD technique [11], the summation of the non-overlapping blocks of size  $8 \times 8$  for  $F_n$  is provided by Equation~(1)

$$\phi_n(x, y) = \frac{1}{w^2} \sum_{u=0}^{w-1} \sum_{v=0}^{w-1} F_n(wx + u, wy + v) \quad (1)$$

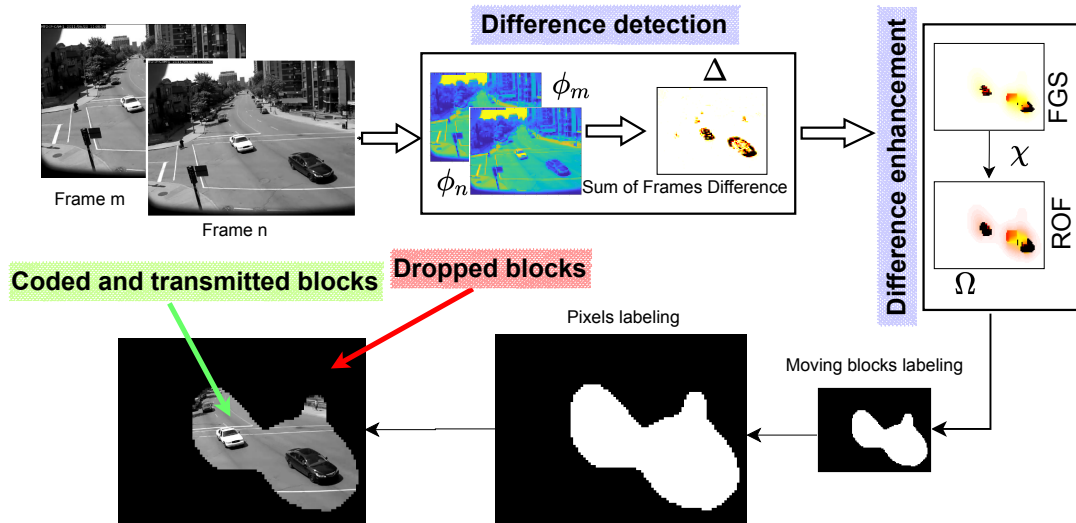


FIGURE 2: Block diagram of the proposed algorithm (BIRD)

While for the frame  $F_m$ ,  $\phi_m$  is calculated using Equation~(2)

$$\phi_m(x, y) = \frac{1}{w^2} \sum_{u=0}^{w-1} \sum_{v=0}^{w-1} F_m(wx + u, wy + v) \quad (2)$$

Where  $x \in 0 \dots M/w - 1$  and  $y \in 0 \dots N/w - 1$  are block indices. The resulting intensity maps  $\phi_n$  and  $\phi_m$  are  $w^2$  times less than the input frame size  $F_n$ . To create the activity map  $\Delta$ , the SFD operation is completed by computing the absolute difference between the two intensity maps, as in Equation~(3)

$$\Delta(w, y) = |\phi_n(x, y) - \phi_m(x, y)| \quad (3)$$

In view of this, the scores in  $\Delta$  indicate the level of activity created between the two frames. The blocks that contain high movement are represented by high score values in  $\Delta$ , which indicates the moving regions. However, lower scores values indicate the non-moving regions. The complete scheme of the proposed method is shown in Figure (2).

### B. DIFFERENCE ENHANCEMENT

To avoid the false negative problem and improve the accuracy, an enhancement of the scores of  $\Delta$  is needed. We propose the combination of a smoothing and rank maximization of  $\Delta$ . Therefore, we propose to take the advantage of both the efficiency and rapidity of the Gaussian smoother the fast global smoother (FGS) [48]. As depicted in Figure (2), FGS is applied on the  $\Delta$  map to smooth the details and noisy part resulting from the SFD operation. Contrary to the convolution filters, FGS is characterized by a low complexity and rapidity estimated to be over 30 times faster than other filters. FGS uses a parameter  $\sigma$  to control the variance around the mean value and another parameter  $\lambda$  to define the amount of regularization during filtering.

### Algorithm 1: The Proposed BIRD algorithm

**Input:**

- $m$  selected previous frame
- $N$  SFD blocks size
- $K$  ROF window size
- $p$  rank order of the ROF
- $T$  threshold value
- $\lambda$  regularization of FGS
- $\sigma$  variance around the mean of FGS

**Output:**

- $Mask$  binary mask of ROI
- $block_{ind}$  vector of ROI blocks indexes

**for Each New frame  $F_n$  do**

**Apply** Equations (1) (2) and (3);

$\Delta \leftarrow SFD(F_n, F_m)$ ;

**Apply** Fast Global Smoother ;

$\chi \leftarrow FGS(\Delta, \lambda, \sigma)$ ;

**Apply** 1-D Rank order filter ;

$\Omega \leftarrow ROF(\chi, K, p)$ ;

**Set**  $T$ ;

**for all scores in  $\Omega$  do**

**if**  $Score(x, y) \geq T$  **then**

**Set**  $mask(block) \leftarrow 1$ ;

**Set**  $block_{ind} \in S_a$  ;

**else**

**Set**  $mask(block) \leftarrow 0$  ;

**end**

**end**

**Report** ROI mask to encoder ;

**Report**  $block_{ind}$  vector to receiver;

**end**

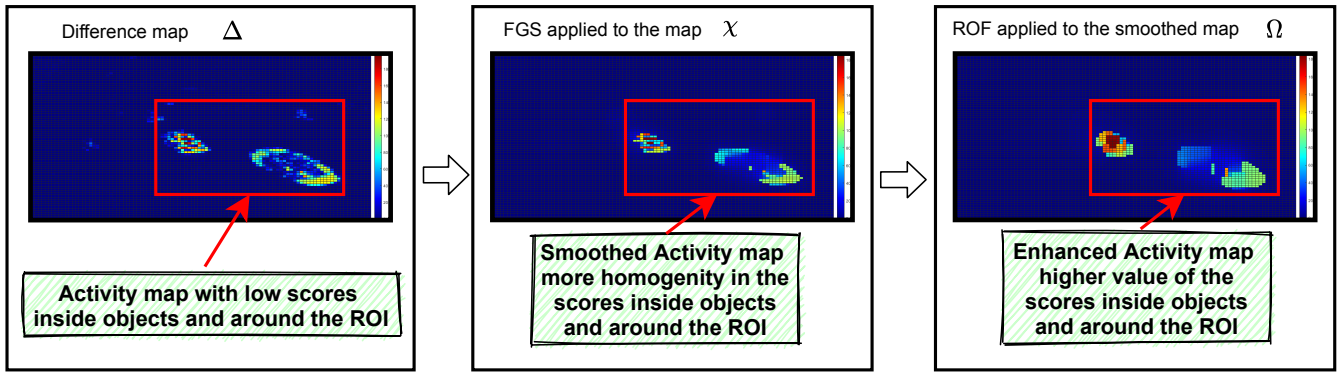


FIGURE 3: Impact of the combination of FGS and ROF on the ROI classification

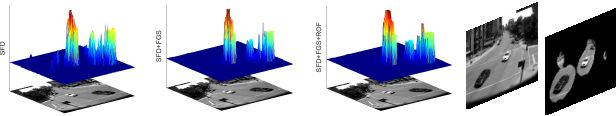


FIGURE 4: FGS eliminates unnecessary activities and ROF enhances the non-zeros scores prior to thresholding

Subsequently, the resulting smoothed map ( $\chi$ ) is filtered by the maximum rank order filter (ROF). The ROF belongs to a class of filters easy to implement [49]. The maximum rank order filter calculates the envelope of the smoothed map. It is a fast and cost-effective solution due to its simple arithmetic operations [23]. Let  $Q = l_1, l_2, \dots, l_k$  be the set of input samples to the filtering process within the predefined observation window. The result of ordering the samples  $l_1, l_2, \dots, l_k$  is obtained by the logical ordering  $l_{(1)}, l_{(2)}, \dots, l_{(N)}$  where  $l_{(i)} \in Q$ , for  $i \in 1 \dots N$  represents the  $i^{th}$  order statistic. The ROF filter uses  $l_{(N)}$  the maximum order statistic. The obtained filtered map is noted  $\Omega$ . Figure (3) illustrates the impact of the used filters to enhance the ROI classification performances while Figure (4) summarizes the impact of each filter as used in this order.

The binary mask is then created by comparing the  $\Omega$  scores to a threshold. Where scores higher than the threshold value indicate activity in the associated block, whereas scores lower than the threshold value indicate inactivity.

Following the threshold operation, a set of block indices ( $S_a$ ) composed of the indexes of the activity blocks is constructed. Based on the proposed strategy, only the ROI blocks will be compressed and sent to the destination. The algorithm 1 further summarizes the above steps.

#### IV. RESULTS AND DISCUSSION

To validate the proposed method, we present the Change Detection 2014 Dataset (CDnet) [50] results. CDnet 2014 is a very challenging dataset composed of 51 video sequences from 11 categories ( more than 150000 frames + their ground truths). Since each category is associated with a specific change detection problem, e.g., dynamic background, shadows, CDnet enables an objective identification and ranking of

methods that are most suitable for a specific problem as well as competent overall. The experimental values for each used parameter are summarized in Table 2.

TABLE 2: Used parameters for the conducted simulations

Step	SFD	FGS	ROF		
Parameter	N	$\sigma$	$\lambda$	$p$	$K$
Value	8	0.05	30	100	4

We consider first a qualitative assessment based on visual observation of the obtained binary mask for the moving regions compared with ground truth masks.

#### A. PARAMETERS AND EVALUATION METRICS

Seven metrics are used for assessment. These are calculated using the confusion matrix that contains the classification characteristics in terms of quality and quantity.

##### 1) Evaluation Metrics

TP: True positives, the number of pixels correctly labeled as foreground.

FP: False positives, the number of pixels incorrectly labeled as foreground.

TN: True negatives, the number of pixels correctly labeled as background.

FN: False negatives, the number of pixels incorrectly set as background.

Seven measures are substituted for the preceding four in order to more accurately assess the classification results. The metrics are given by Equations (4)-(11).

Recall:

$$Re = \frac{TP}{TP + FN} \quad (4)$$

Specificity:

$$Sp = \frac{TN}{TN + FP} \quad (5)$$

Precision:

$$Pr = \frac{TP}{TP + FP} \quad (6)$$

F-measure:

$$Fm = 2 \frac{Pr}{Re + Pr} \quad (7)$$

False-positive rate (FPR):

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

False-negative rate (FNR):

$$FNR = \frac{FN}{TP + FN} \quad (9)$$

Percentage of wrong classifications (PWC):

$$PWC = 100 \frac{(FN + FP)}{(TP + FN + FP + TN)} \quad (10)$$

Balanced-Accuracy (BAC):

$$BAC = \frac{Re + Sp}{2} \quad (11)$$

For PWC, FNR, and FPR metrics, lower values indicate higher accuracy, but for Recall, Specificity, Precision, BAC and F-Measure, higher values indicate better performance [35]. Recall gives the percentage of necessary positives via the compared total number of true positive pixels in the ground truth. Precision gives the percentage of unnecessary positives through the compared total number of positive pixels in the detected binary objects mask.

Among these metrics, we are specifically interested in the recall and balanced-Accuracy metrics (BAC). ROI-based video coding needs a high TP with a minimum FN.

Advanced analysis is performed by exposing the TPR-FPR curve (ROC curve) for sample sequences with an analysis of the optimum threshold.

### B. PERFORMANCES OF BIRD OVER THE CDNET 2014

Table 3 shows the performance of BIRD indicating the algorithm's visual accuracy in detecting all the ROI candidates for compression and transmission. The presented sample frames from all categories of the benchmark dataset in Table 3 show that the algorithm successfully detects the blocks in which a high movement occurs. Objects are entirely detected in most videos, which could be a good enabler for a variety of applications, especially as a pre-encoder for ROI-based video coding [23].

It should be noted that, for some video scenarios (like the *Office* video sample), the algorithm is unable to detect the target object for some time due to the object's stability. Even though the object information has already been delivered to the destination, the reported numerical results are reduced.

Table 4 shows the quantitative results on CDnet 2014 dataset. The results indicate the good performance of the proposed algorithm in the detection of the whole object with high TP values for different categories. The algorithm shows high detection results for some categories and moderate detection performances for others. For example, the recall metric is high for almost all the categories but shows exceptional performance for night video and dynamic background, PTZ,

and camera jitter categories despite their difficult scenarios. The algorithm presents some weaknesses in detecting the complete object in some categories like intermittent object motion category.

### C. COMPARISON WITH OTHER TECHNIQUES

Table 5 shows the overall results of our method on CDnet 2014 dataset compared with the state-of-the-art techniques namely, KNN in [51], GMM in [52], KDE in [53], Mahalanobis Distance and Euclidean Distance techniques presented in [54] and another GMM-based technique in [55]. The proposed method exhibits good results in the recall and FNR metrics with the best results against other techniques and shows competitive results for the specificity metric.

The weaknesses of the algorithm in the precision and F-measure values (0.1893 and 0.2678) can be explained by the adopted block-based techniques which allow the detection of additional pixels with the moving object, (i.e.: high FPR).

According to Table 4, the results of BIRD are considered very high in the context of the studies that aim to integrate object detection as a pre-processing step for WVS in very low-complexity platforms.

### D. METRICS OF INTEREST: RECALL, SPECIFICITY AND BAC

A balance between the TP and FN is important to measure the performance of BIRD in detecting the complete object while avoiding the drawback of the non-detection of regions inside the moving objects and with the minimum FP possible. We compare BIRD to two methods, one method uses Neural Networks for object detection [24]. The second method uses block-based object detection [56] same as our proposed method.

As presented in Table 6, the BAC and recall metrics of BIRD show higher values than in [56] for most of the sequences. While [24] shows superior BAC and specificity values compared with BIRD and [56]. Results of BIRD are still very competitive to that of [24]. With an overall BAC of 82%, BIRD can ensure high detection accuracy of the moving object regions for different categories and conditions.





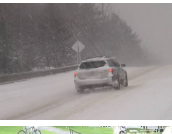


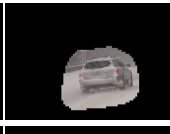




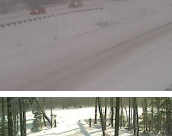


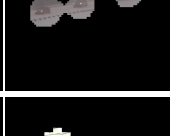








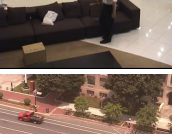


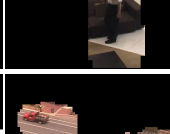

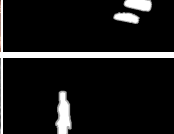

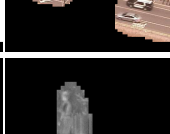




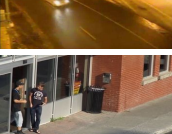


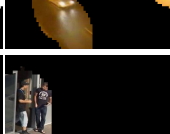







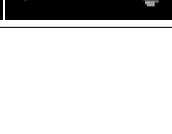
### E. THE IMPACT OF THRESHOLDING ON DETECTION

We select three sequences from the used dataset to empirically validate the BIRD accuracy and low-overhead assumptions. *Highway* with a size of (320 × 240) contains high activity with a number of moving vehicles. The *pedestrians* sequence of size (360 × 240) is of low activity with relatively high stability in the background. The *Snowfall* sequence of size (720 × 480) is a long sequence that contains moving objects with very high activity in the background (Snow and winter).

Figure (5) plots the TPR against the FPR when varying the threshold value (0...10). The obtained ROC curves show that low thresholds imply a high true positive rate. However, this adversely affects the specificity of the detection, since a high number of blocks is wrongly labeled as activity blocks,



TABLE 3: Samples of ROI extraction mask results

Sequence	Original	ground-truth	mask	ROI
Highway #1475				
SnowFall #2784				
Pedestrians #476				
Blizzard #1406				
WinterDriveway #1860				
tunnelExit #2329				
Sofa #1185				
PTZ #1240				
Park #250				
NightVideo #1300				
Busstation #400				
Turbulance0 #2045				

Category	Recall	Specificity	FPR	FNR	PBC	Precision	F-Measure
<b>PTZ</b>	0.9662	0.6443	0.3556	0.0337	35.3016	0.0401	0.0753
<b>badWeat.</b>	0.9208	0.8948	0.1051	0.0791	10.1795	0.2747	0.3904
<b>baseline</b>	0.7619	0.9437	0.0562	0.2380	6.6360	0.3268	0.4047
<b>cameraJ.</b>	0.8504	0.6446	0.3553	0.1495	34.5590	0.1383	0.2238
<b>dynamic.</b>	0.7593	0.9512	0.0487	0.2406	4.9399	0.1962	0.2801
<b>intermi.</b>	0.4186	0.8603	0.1396	0.5813	16.4228	0.1566	0.2242
<b>lowFram.</b>	0.8161	0.7905	0.2094	0.1838	20.2242	0.1315	0.1919
<b>nightVi.</b>	0.9455	0.8374	0.1625	0.0544	15.9206	0.1193	0.2108
<b>shadow</b>	0.8775	0.8500	0.1499	0.1224	14.8039	0.2416	0.3740
<b>thermal</b>	0.7548	0.8894	0.1105	0.2451	13.4618	0.3575	0.4095
<b>turbule.</b>	0.8216	0.8870	0.1129	0.1783	11.3767	0.1000	0.1607
<b>Overall</b>	<b>0.8084</b>	<b>0.8357</b>	<b>0.1642</b>	<b>0.1915</b>	<b>16.7115</b>	<b>0.1893</b>	<b>0.2678</b>

TABLE 4: Detection results of the proposed algorithm over CDnet 2014 dataset

Technique	Recall	Specificity	FPR	FNR	PWC	F-Measure	Precision
<b>KNN [51]</b>	0.6650	0.9802	0.0198	0.3350	<b>3.3200</b>	<b>0.5937</b>	0.6788
<b>GMM1 [52]</b>	0.6846	0.9750	0.0250	0.3154	3.7667	0.5707	0.6025
<b>KDE [53]</b>	0.7375	0.9519	0.0481	0.2625	5.6262	0.5688	0.5811
<b>MahaD [54]</b>	0.1644	<b>0.9931</b>	<b>0.0069</b>	0.8356	3.4750	0.2267	<b>0.7403</b>
<b>GMM2 [55]</b>	0.6604	0.9725	0.0275	0.3396	3.9953	0.5566	0.5973
<b>EucD [54]</b>	0.6803	0.9449	0.0551	0.3197	6.5423	0.5161	0.5480
<b>BIRD</b>	<b>0.8084</b>	0.8357	0.1642	<b>0.1915</b>	16.7115	0.1893	0.2678

TABLE 5: Comparison of BIRD with classical techniques over CDnet 2014 dataset

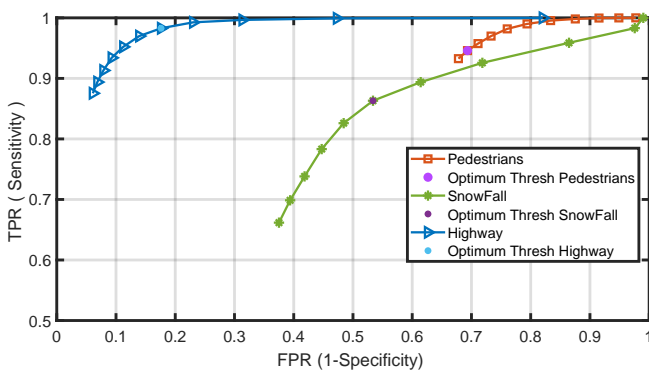


FIGURE 5: ROC curve and the optimum threshold for *pedestrians*, *Highway* and *Snowfall* sequences

which means that more data is to be considered for delivery. The optimum threshold that allows the best tradeoff between TPR and FPR could be achieved as shown by the orange dots in each ROC curve. It is defined by calculating the minimum Gaussian distance between the results of TPR and FPR:  $\min(\sqrt{(1 - sensitivity)^2 + (specificity - 1)^2})$ .

Figure (6) shows the impact of varying the threshold value on the mean value of the detected blocks. In the case where high stability characterizes the background (for example *pedestrians* sequence), a high threshold is generally preferred since there is a low risk of wrongly including background blocks in the ROI. Meanwhile, a high number of background blocks is classified as ROI in the case of noisy and dynamic background (the Snowing scene in the *Snowfall* sequence for example). A higher number of the ROI detected blocks may enhance the quality of the reconstructed frames at the

TABLE 6: Category-wise comparison of BIRD with the state-of-the-art on CDnet 2014 dataset

Category	Recall			Specificity			Balanced Acc.		
	BIRD	Savas [56]	Cwizar [24]	BIRD	Savas [56]	Cwizar [24]	BIRD	Savas [56]	Cwizar [24]
Dynamic.	0.7593	0.6436	<b>0.8144</b>	0.9512	0.9962	<b>0.9985</b>	0.8553	0.8199	<b>0.9064</b>
PTZ	<b>0.9662</b>	0.7685	0.3833	0.6443	<b>0.9977</b>	0.9968	0.8053	<b>0.8831</b>	0.6901
BadWeat.	<b>0.9208</b>	0.5647	0.6697	0.8948	0.9985	<b>0.9993</b>	<b>0.9078</b>	0.7816	0.8345
Baseline	0.7619	0.6214	<b>0.8972</b>	0.9437	0.8213	<b>0.9980</b>	0.8528	0.7213	<b>0.9476</b>
CameraJ.	<b>0.8504</b>	0.4567	0.7436	0.6446	0.9788	<b>0.9931</b>	0.7475	0.7177	<b>0.8683</b>
Intermi.	0.4186	0.5547	<b>0.8324</b>	0.8603	<b>0.9979</b>	0.9911	0.6394	0.7763	<b>0.9118</b>
LowFram.	<b>0.8161</b>	0.5490	0.6659	0.7905	0.7464	<b>0.9949</b>	0.8033	0.6477	<b>0.8304</b>
nightVi.	<b>0.9455</b>	0.4593	0.4511	0.8374	0.9583	<b>0.9874</b>	<b>0.8915</b>	0.7088	0.7193
Shadow	0.8775	0.8365	<b>0.8786</b>	0.8500	0.9828	<b>0.9910</b>	0.8638	0.9097	<b>0.9348</b>
Thermal	<b>0.7548</b>	0.4650	0.7268	0.8894	0.9647	<b>0.9949</b>	0.8221	0.7148	<b>0.8609</b>
Turbule.	<b>0.8216</b>	0.7421	0.7122	0.8870	0.9883	<b>0.9997</b>	0.8543	<b>0.8652</b>	0.8559
Overall	<b>0.8084</b>	0.6056	<b>0.6608</b>	0.8357	<b>0.9483</b>	<b>0.9948</b>	<b>0.8220</b>	0.7770	<b>0.8509</b>

\***bold** values are the best category-wise, **red** values are the best overall, **blue** values are the second best

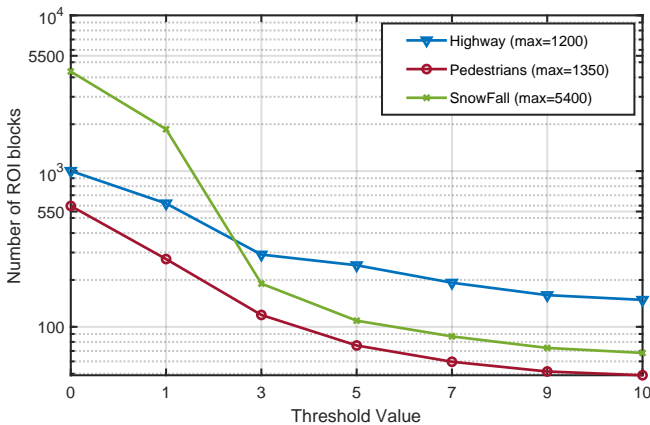


FIGURE 6: Number of blocks belonging to the ROI according to the threshold value

destination. But, at the cost of higher energy and bitrate.

Table 7 shows the impact of the threshold value on the energy gain expressed by the number of skipped blocks. From the table, it can be seen that the mean number of ROI blocks is inversely proportional to the threshold value. As a result, the energy gain is low when the chosen threshold value is low. A borderline case is when the threshold value is 0 (i.e.the activity score is absolutely greater than 0), which gives the lowest energy gain. The row that begins with MAX, indicates that all the frame's blocks will be compressed and

transmitted (i.e.including the blocks in which the activity score is equal to 0). In this case, all the frame's blocks are taken into account for compression and transmission, rendering the method ineffective. According to the accuracy results shown in Figure 5, for the *pedestrians* sequence, the optimum threshold for good detection accuracy is 9. Consequently, this threshold value enables a saving of about 96% of the processing and transmission energy compared to the CTA approach (Table 7). Choosing a low threshold value is without benefit to the surveillance system, while an optimum threshold could significantly save the energy consumption in the sensor node and the bitrate needed for transmission. Furthermore, an optimum threshold enables the optimum ratio of the activity blocks and could be used as a rate controller, which is an interesting subject for future work.

#### F. METHOD COMPLEXITY

To evaluate the consumed energy on embedded sensor conditions, we have considered what follows, a sensor node equipped with an ARM Cortex M3 micro-controller [57]. Table 8 shows the processor characteristics. Using MATLAB 2020a and C++ running on a PC intel Core i7-2670QM 2.2Ghz, with 8GB RAM on Windows 7 OS, 2.6 ms to process one frame of 320x240 is recorded allowing processing of 384 frames per second (fps).

TABLE 7: Statistics of the energy gain under threshold variation

Threshold	Highway			Pedestrians			Snowfall		
	mean (ROI) <sub>(ceiled)</sub>	ratio (ROI)	$\Delta$ energy (theoretically)	mean (ROI)	ratio (ROI)	$\Delta$ energy (theoretically)	mean (ROI)	ratio (ROI)	$\Delta$ energy (theoretically)
10	149	12.41%	<b>+87.59%</b>	49	03.63%	<b>+96.37%</b>	68	01.26%	<b>+98.74%</b>
9	160	13.33%	<b>+86.67%</b>	52	03.85%	<b>+96.15%</b>	74	01.37%	<b>+98.63%</b>
7	192	16.00%	<b>+84.00%</b>	60	04.44%	<b>+95.56%</b>	87	01.61%	<b>+98.39%</b>
5	249	20.75%	<b>+79.25%</b>	76	05.63%	<b>+94.37%</b>	110	02.04%	<b>+97.96%</b>
3	291	24.25%	<b>+75.75%</b>	120	08.89%	<b>+91.11%</b>	190	03.52%	<b>+96.48%</b>
1	621	51.75%	<b>+48.25%</b>	273	20.22%	<b>+79.78%</b>	1857	34.39%	<b>+65.61%</b>
0	1003	83.58%	<b>+16.42%</b>	598	44.30%	<b>+55.70%</b>	4360	80.74%	<b>+19.26%</b>
Max	1200	100%	-	1350	100%	-	5400	100%	-

Sensor Processor	Cortex M3
Clock rate	72 MHz
Processor power	23 mW
Cycles count	Add.[1], Sub.[1], Mult.[1 or 2], Div.[1 to 12].

TABLE 8: ARM Cortex M3 characteristics

1) Energy Budget for change detection

The total energy budget of the proposed BIRD algorithm is directly proportional to its computational complexity and could be expressed as follows:

$$E_{Detection} = E_{SFD} + E_{FGS} + E_{ROF} + E_{Threshold} \quad (12)$$

The total computational budget of the method is presented in Table 9. The number of operations for FGS is reported in [48], while the ROF budget is  $R = K(K - 1)/2$ , where  $K$  is set to 4 for the proposed method and represents the size of the sliding vector. The filter uses the sliding vector over the columns. After each calculation step, the vector is shifted by one position down, and the operation is executed till the end of the line vector. This process is performed along all the columns. For  $K = 4$ , the ROF performs 6 comparisons for each score value in the map.

Since the number of operations performed is proportional to the frame size and the block size ( $8 \times 8, 16 \times 16 \dots$ ), a generalized model of the number of arithmetic operations should be presented. We present in Table 9 the number of operations for each step in terms of frame size ( $N, M$ ) and block size ( $w$ ). Table 9 also shows the energy budget of each step and the total energy budget of the BIRD. Table 10 shows a comparison of the energy budget of the proposed object detection method against state-of-the-art techniques for  $240 \times 320$ , namely MoG [52], CS-MoC [58], CoSCS-MoG [59], EBSCAM [60] and the basic FD technique. The

proposed technique shows the lowest energy consumption records in both its minimal and maximal cases. While energy consumption recorded an increase of about 38% compared to FD when extreme cases are considered.

2) Energy dissipation for complete compression chain

Considering a complete compression chain, the total in-node processing budget could be expressed as follow:

$$E_{total} = E_{Detection} + E_{compress} \quad (13)$$

Where  $E_{Detection}$  is the energy cost of the object detection part as presented by Equation (12),  $E_{compress}$  is the energy cost of the compression part. For the calculation of  $E_{compress}$ , the model has been studied and provided in [61] under the same conditions.

The compression cost for each frame includes the DCT compression, the quantization cost and the Huffman coding cost. Three implementations of the JPEG-based compression are shown in [61] namely float IJG, slow IJG and fast IJG. In this work, the slow IJG implementation is adopted with an energy cost of  $192.28 \mu J$  for each  $8 \times 8$  block.

Since  $N_{blocks}$  represents the number of activity blocks detected that will be coded for each frame, the compression cost is proportionally related to  $N_{blocks}$ . For example, the Highway sequence records an overhead of the object detection step  $E_{Detection}$  equal to  $0.6891 mJ/frame$ .

Figure (7) illustrates the per-frame energy consumption of the proposed method compared to ROI-based compression methods, namely, [30] referred to as EMP'22, [16] referred to as SSD'22 and the forward baseline compression (MJPEG). Since the algorithm is applied to each frame, constant energy is spent for each frame, while the total energy curves oscillate based on the number of blocks to compress. BIRD shows the best results as the lowest energy budget for all the scenarios.

TABLE 9: Computational budget of each step of BIRD algorithm

Step	Operations	# of Operations	Energy consumption (mJ/Frame)	
			min ( $Cycles_{div} = 1$ )	max ( $Cycles_{div} = 12$ )
SFD	Addition	$NM - NM/w^2$	0.2693	0.4
	Subtraction	$NM/w^2$		
	Absolute	$NM/w^2$		
	Division	$NM/w^2$		
ROF	Comparison	$6(N/w^2 - 3)M/w^2$	$7.4250e^{-5}$	$7.4250e^{-5}$
FGS	Multiplication	$6NM/w^2$	0.0832	0.2851
	Division	$NM/w^2$		
Thresholding	Comparison	$NM/w^2$	0.004	0.004
$E_{detection}$	-	-	<b>0.3723</b>	<b>0.6891</b>

TABLE 10: Per-frame  $E_{detection}$  cost of the method compared to state-of-the-art for size  $(240 \times 320)$

Method	Energy Budget (mJ/Frame)	
	min ( $Cycles_{div} = 1$ )	max ( $Cycles_{div} = 12$ )
MoG [52]		649.95
CS-MoG [58]		116.44
CoSCS-MoG [59]		125.96
EBSCAM [60]		3.4
FD		0.5069
BIRD (proposed)	0.3723	0.6891

The energy dissipation of the BIRD method is proportional to the frame size. About 79.29% of blocks are skipped for the *Highway* sequence compared to the standard coding (MJPEG for example), while more than 98% of the blocks are skipped for *SnowFall* sequence and 86.89% for *pedestrians* sequence. The level of energy consumption at the processing step is correlated with the number of skipped blocks.

Despite the good ROI detection of the other techniques, they are weakened by the high energy cost in the detection step. This is due to the adopted edge detection and automatic thresholding techniques in [30] [16] respectively. Those techniques are computationally extensive due to the use of arithmetic convolution and histogram calculation. Meanwhile, the optimized design of edge detectors and otsu's threshold should help reduce their energy budget.

From Figure (7) we can deduce that the algorithm is

efficient in saving a substantial amount of processing and transmission power. The saving achieves more than 90% of the energy most of the time. The proposed method provides a good balance between energy saving and detection accuracy.

### 3) Memory requirements

We analyze here the memory requirement of the proposed region detection method. The method requires storing the previous grayscale frame of 8-bit depth and updating every frame, corresponding to a memory of  $N \times M$  bytes. Two score maps are to be stored which requires a memory of  $2 \times N \times M/w^2$  bytes. The ROF and the FGS filters are performed locally on the stored activity map. Thus, the needed memory for these operations is ignored (window of 4 Bytes for ROF and short vectors for FGS). For  $w = 8$ , the total memory consumption is about 1.031 bytes per pixel.

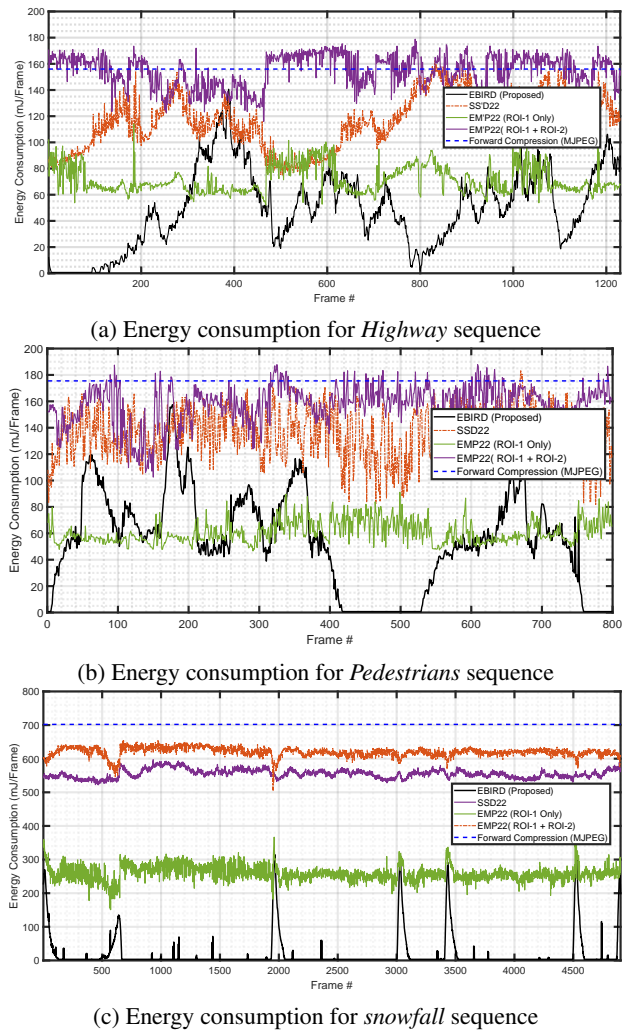


FIGURE 7: Per-frame energy dissipation of BIRD for *Highway*, *pedestrians* and *Snowfall*

## V. CONCLUSION

In this study, we proposed an energy-efficient moving region detection approach as a pre-encoder for WVS. The suggested approach is built upon a low-complexity SFD operation followed by morphological filtering and thresholding. The proposed method's overall efficiency was evaluated using a standard dataset as a benchmark. The performance assessment shows a satisfactory balance between the proposed method's detection accuracy, energy efficiency, and memory. In these respects, our approach effectively relieves the burden of processing and compressing video sequences for resource-constrained surveillance devices. The proposed method has two main drawbacks: (1) It has only been tested on fixed cameras, and (2) It has, in some cases, poor results using some performance metrics (like F-measure) because of its commitment to meet constrained WVS. Future studies include improving the performance of the algorithm and its implementation in an embedded WVS system while taking into account channel and network characteristics.

## REFERENCES

- [1] S. H. Abdhussain, A. R. Ramli, M. I. Saripan, B. M. Mahmmod, S. A. R. Al-Haddad, and W. A. Jassim, "Methods and challenges in shot boundary detection: a review," *Entropy*, vol. 20, no. 4, p. 214, 2018.
- [2] A. Redondi, L. Baroffio, L. Bianchi, M. Cesana, and M. Tagliasacchi, "Compress-then-analyze versus analyze-then-compress: What is best in visual sensor networks?" *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 3000–3013, 2016.
- [3] A. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Compress-then-analyze vs. analyze-then-compress: Two paradigms for image analysis in visual sensor networks," in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSp)*. IEEE, 2013, pp. 278–282.
- [4] I. Mansri, N. Doghmane, N. Kouadria, S. Harize, and A. Bekhouch, "Comparative evaluation of VVC, HEVC, H. 264, AV1, and VP9 encoders for low-delay video applications," in *2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA)*. IEEE, 2020, pp. 38–43.
- [5] S. Fossel, G. Fottinger, and J. Mohr, "Motion JPEG2000 for high quality video systems," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 787–791, 2003.
- [6] I. Telecom et al., "Advanced video coding for generic audiovisual services," ITU-T Recommendation H. 264, 2003.
- [7] S. Harize, A. Mefoued, N. Kouadria, and N. Doghmane, "HEVC transforms with reduced elements bit depth," *Electronics Letters*, vol. 54, no. 22, pp. 1278–1280, 2018.
- [8] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, "Coding visual features extracted from video sequences," *IEEE transactions on Image Processing*, vol. 23, no. 5, pp. 2262–2276, 2014.
- [9] A. Boulmaiz, N. Doghmane, S. Harize, N. Kouadria, and D. Messadeg, "The use of WSN (wireless sensor network) in the surveillance of endangered bird species," in *Advances in ubiquitous computing*. Elsevier, 2020, pp. 261–306.
- [10] A. Sakhri, O. Hadji, C. Bouarrouguen, M. Maimour, N. Kouadria, A. Benyahia, E. Rondeau, N. Doghmane, and S. Harize, "Audio-visual low power system for endangered waterbirds monitoring," *IFAC-PapersOnLine*, vol. 55, no. 5, pp. 25–30, 2022.
- [11] L. Kong and R. Dai, "Efficient video encoding for automatic video analysis in distributed wireless surveillance systems," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 3, pp. 1–24, 2018.
- [12] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in multimedia*, vol. 2009, 2009.
- [13] Z. Pan, L. Chen, and X. Sun, "Low complexity HEVC encoder for visual sensor networks," *Sensors*, vol. 15, no. 12, pp. 30 115–30 125, 2015.
- [14] X. Jiang, J. Feng, T. Song, and T. Katayama, "Low-complexity and hardware-friendly H. 265/HEVC encoder for vehicular ad-hoc networks," *Sensors*, vol. 19, no. 8, p. 1927, 2019.
- [15] M. Maimour, "SenseVid: A traffic trace based tool for QoE Video transmission assessment dedicated to Wireless Video Sensor Networks," *Simulation Modelling Practice and Theory*, vol. 87, pp. 120–137, 2018.
- [16] A. Aliouat, N. Kouadria, M. Maimour, and S. Harize, "Region-of-interest based video coding strategy for low bitrate surveillance systems," in *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2022, pp. 1357–1362.
- [17] M. A. Hossain, M. I. Hossain, M. D. Hossain, and E.-N. Huh, "DFC-D: A dynamic weight-based multiple features combination for real-time moving object detection," *Multimedia Tools and Applications*, pp. 1–32, 2022.
- [18] H. Wei and Q. Peng, "A block-wise frame difference method for real-time video motion detection," *International Journal of Advanced Robotic Systems*, vol. 15, no. 4, p. 1729881418783633, 2018.
- [19] B. Laugraud, S. Piérard, M. Braham, and M. Van Droogenbroeck, "Simple median-based method for stationary background generation using background subtraction algorithms," in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 477–484.
- [20] L. Kong and R. Dai, "Object-detection-based video compression for wireless surveillance systems," *IEEE MultiMedia*, vol. 24, no. 2, pp. 76–85, 2017.
- [21] L. Galteri, M. Bertini, L. Seidenari, and A. Del Bimbo, "Video compression for object detection algorithms," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3007–3012.
- [22] N. Kouadria, K. Mechouek, S. Harize, and N. Doghmane, "Region-of-interest based image compression using the discrete tchebichef transform in wireless visual sensor networks," *Computers & Electrical Engineering*, vol. 73, pp. 194–208, 2019.

- [23] J. H. Ko, T. Na, and S. Mukhopadhyay, "An energy-quality scalable wireless image sensor node for object-based video surveillance," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 3, pp. 591–602, 2018.
- [24] M. De Gregorio and M. Giordano, "Change detection with weightless neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 403–407.
- [25] W. Guo, C. Yan, and T. Lu, "Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing," *International Journal of Distributed Sensor Networks*, vol. 15, no. 2, p. 1550147719833541, 2019.
- [26] S. M. Chowdhury and A. Hossain, "Different energy saving schemes in wireless sensor networks: A survey," *Wireless Personal Communications*, vol. 114, no. 3, pp. 2043–2062, 2020.
- [27] D. K. Sah and T. Amgoth, "Parametric survey on cross-layer designs for wireless sensor networks," *Computer Science Review*, vol. 27, pp. 112–134, 2018.
- [28] A. A. Youssif, A. Z. Ghalwash *et al.*, "Energy aware and adaptive cross layer scheme for video transmission over wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 21, pp. 7792–7802, 2016.
- [29] Y. A. U. Rehman, M. Tariq, and T. Sato, "A novel energy efficient object detection and image transmission approach for wireless multimedia sensor networks," *IEEE sensors journal*, vol. 16, no. 15, pp. 5942–5949, 2016.
- [30] A. Aliouat, N. Kouadria, S. Harize, and M. Maimour, "Multi-Threshold-Based Frame Segmentation for Content-Aware Video Coding in WMSN," in *International Conference on Computing Systems and Applications*. Springer, 2022, pp. 337–347.
- [31] J. H. Ko, B. A. Mudassar, and S. Mukhopadhyay, "An energy-efficient wireless video sensor node for moving object surveillance," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 1, no. 1, pp. 7–18, 2015.
- [32] A. Aliouat, N. Kouadria, M. Maimour, S. Harize, and N. Doghmane, "Region-of-interest based video coding strategy for rate/energy-constrained smart surveillance systems using wmsns," *Ad Hoc Networks*, p. 103076, 2022.
- [33] S. S. Sengar and S. Mukhopadhyay, "Motion segmentation-based surveillance video compression using adaptive particle swarm optimization," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11 443–11 457, 2020.
- [34] O. Iqbal, V. I. T. Muro, S. Katoch, A. Spanias, and S. Jayasuriya, "Adaptive subsampling for roi-based visual tracking: Algorithms and fpga implementation," *IEEE Access*, vol. 10, pp. 90 507–90 522, 2022.
- [35] I. M. Hameed, S. H. Abdullhussain, and B. M. Mahmmod, "Content-based image retrieval: A review of recent trends," *Cogent Engineering*, vol. 8, no. 1, p. 1927469, 2021.
- [36] M. Fularz, M. Kraft, A. Schmidt, and A. Kasiński, "A high-performance FPGA-based image feature detector and matcher based on the FAST and BRIEF algorithms," *International Journal of Advanced Robotic Systems*, vol. 12, no. 10, p. 141, 2015.
- [37] P. Kumar, A. Singhal, S. Mehta, and A. Mittal, "Real-time moving object detection algorithm on high-resolution videos using GPUs," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 93–109, 2016.
- [38] K. Goyal and J. Singhai, "Review of background subtraction methods using gaussian mixture model for video surveillance systems," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 241–259, 2018.
- [39] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-time imaging*, vol. 11, no. 3, pp. 172–185, 2005.
- [40] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image processing*, vol. 20, no. 6, pp. 1709–1724, 2010.
- [41] R. Antonio, S. Faria, L. M. Tavora, A. Navarro, and P. Assuncao, "Learning-based compression of visual objects for smart surveillance," in *2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2022, pp. 1–6.
- [42] H. Zhu, X. Yan, H. Tang, Y. Chang, B. Li, and X. Yuan, "Moving object detection with deep CNNs," *IEEE Access*, vol. 8, pp. 29 729–29 741, 2020.
- [43] S. S. Sengar and S. Mukhopadhyay, "Moving object detection based on frame difference and W4," *Signal, Image and Video Processing*, vol. 11, no. 7, pp. 1357–1364, 2017.
- [44] S. H. Shaikh, K. Saeed, and N. Chaki, "Moving object detection using background subtraction," in *Moving object detection using background subtraction*. Springer, 2014, pp. 15–23.
- [45] A. Haidous, W. Oswald, H. Das, and N. Gong, "Content-adaptable roi-aware video storage for power-quality scalable mobile streaming," *IEEE Access*, vol. 10, pp. 26 830–26 848, 2022.
- [46] B. Li, L. Ye, J. Liang, Y. Wang, and J. Han, "Region-of-interest and channel attention-based joint optimization of image compression and computer vision," *Neurocomputing*, 2022.
- [47] G. Wu, M. Qin, T. M. Bae, S. Li, Y. Fang, and Y.-K. Chen, "Region of interest quality controllable video coding techniques," Mar. 15 2022, *US Patent 11,277,626*.
- [48] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5638–5653, 2014.
- [49] N. R. Harvey and S. Marshall, "Rank-order morphological filters: A new class of filters," in *IEEE Workshop on nonlinear signal and image processing*. Citeseer, 1995, pp. 975–978.
- [50] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnets 2014: An expanded change detection benchmark dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 387–394.
- [51] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773–780, 2006.
- [52] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, vol. 2. IEEE, 1999, pp. 246–252.
- [53] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *European conference on computer vision*. Springer, 2000, pp. 751–767.
- [54] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative study of background subtraction algorithms," *Journal of Electronic Imaging*, vol. 19, no. 3, p. 033003, 2010.
- [55] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2. IEEE, 2004, pp. 28–31.
- [56] M. F. Savaş, H. Demirel, and B. Erkal, "Moving object detection using an adaptive background subtraction method based on block-based structure in dynamic scene," *Optik*, vol. 168, pp. 605–618, 2018.
- [57] Arm, "Cortex M3 datasheet," <https://iot-lab.github.io/assets/misc/docs/iot-lab-m3/stm32f103re.pdf>, 2018.
- [58] Y. Shen, W. Hu, J. Liu, M. Yang, B. Wei, and C. T. Chou, "Efficient background subtraction for real-time tracking in embedded camera networks," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, 2012, pp. 295–308.
- [59] Y. Shen, W. Hu, M. Yang, J. Liu, B. Wei, S. Lucey, and C. T. Chou, "Real-time and robust compressive background subtraction for embedded camera networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, pp. 406–418, 2015.
- [60] M. U. K. Khan, A. Khan, and C.-M. Kyung, "EBSCam: Background subtraction for ubiquitous computing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 1, pp. 35–47, 2016.
- [61] D.-U. Lee, H. Kim, M. Rahimi, D. Estrin, and J. D. Villasenor, "Energy-efficient image compression for resource-constrained platforms," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 2100–2113, 2009.



**AHGEN ALIOUAT** Is an IEEE student member (97959071), he received his bachelor's and master's degrees in 2016 and 2018, successively in telecommunication, networks, and multimedia from the telecommunication department of the University of Sciences and Technology Houari Boumediene (USTHB), Algeria. Currently enrolled as a Ph.D. student in Multimedia and Digital Communications at the electronics department of Badji Mokhtar University of Annaba (UBMA) in

Algeria since 2019. His main research interests include image and video coding, signal processing, digital communications, and artificial intelligence. His Ph.D. work interests include video/image coding and transmission over wireless networks with resource constraints.



**NASREDDINE KOUADRIA** Received his Ph.D. and the habilitation to direct research in Multimedia and digital communication from Badji Mokhtar University of Annaba, Algeria in 2014 and 2018, respectively. He was a postdoctoral research associate with the technical university of Iasi - Romania in 2019. Currently, he is a senior lecturer in the department of electronics at Annaba university and a research member of the laboratory of automation and signals of Annaba (LASA). His

research interests include WSNs, image/video coding, fast transformations, communication, and digital signal processing.



**SALIHA HARIZE** : Received a Ph.D. in 2014 and a habilitation to direct research in 2017 from the University of Badji Mokhtar-Annaba, Algeria. She is a senior lecturer and member of the laboratory of automatic and signal processing of Annaba (LASA). Her interests include image and video compression, and the implementation of digital circuits on FPGA and cryptography.



**MOUFIDA MAIMOUR** Is an Associate Professor at Lorraine University in Nancy, France, member of the CRAN (Centre de Recherches en Automatique de Nancy) since 2004. She received her engineer degree in Computer Science from the University of Constantine (Algeria) in 1998. She conducted her doctoral research at the LIP (Laboratoire Informatique du Parallélisme) at the ENS of Lyon, France and obtained her PhD from Claude Bernard University in 2003. She mainly

worked on reliable multicast and congestion control in the Internet and took part in multiple networking and grid computing projects. Her current topics include wireless sensor networks, knowledge-defined networks and digital twinning for the IoT with a focus on the IIoT (Industrial Internet of Things) and the IoMT (Internet of Multimedia Things). Machine learning algorithms are leveraged to solve issues related to Industry 4.0 in the former and to multimedia environment surveillance in the latter.

...