

Ministry of Higher Education and Scientific Research

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculty of Technology



Department of Electronics

جامعة باجي مختار – عنابة

كلية التكنولوجيا

قسم الإلكترونيك

Thesis

Submitted to obtain the diploma of

Doctorate Third Cycle

Field : Telecommunication

Specialty : Multimedia and Digital Communications

By :

ALIOUAT Ahcen

Title :

**Etude et mise en œuvre d'un encodeur vidéo basé objet
pour les systèmes de vidéosurveillance sans fils
embarqués**

Thesis defended on 31/05/2023 in front of the jury composed of:

N°	Name and Surname	Grade	Establishment	Role
01	BENOUARET Mohamed	Prof.	Badji Mokhtar – Annaba University	President
02	KOUADRIA Nasreddine	MCA	Badji Mokhtar – Annaba University	Supervisor
03	HARIZE Saliha	MCA	Badji Mokhtar – Annaba University	Co-Supervisor
04	BOUMEHREZ Farouk	MCA	University of Khenchela – Khenchela	Examiner
05	BOUKAACHE Abdenour	MCA	University 8 mai 1945 – Guelma	Examiner
06	MAIMOUR Moufida	MC	Lorraine University – Nancy, France	Invited

"دراسة وتنفيذ برنامج ترميز للفيديو المعتمد على الكائنات لأنظمة المراقبة بالفيديو الدمجة اللاسلكية "

الملخص:

تكتسب أنظمة المراقبة بالفيديو اللاسلكية المدمجة شعبية واسعة بسبب التقدم في أنظمة إنترنت الأشياء (IoT) وشبكات استشعار الوسائط المتعددة اللاسلكية. هذه الأنظمة لها العديد من التطبيقات ، بما في ذلك تتبع الأهداف العسكرية والمراقبة ، والإغاثة في حالات الكوارث ، ومراقبة الصحة الطبية الحيوية ، والاستشعار الزلزالي ، ومراقبة البيئة ، والمدن الذكية. لا يزال نقل بيانات الوسائط المتعددة بمعدل بت منخفض مع الحفاظ على البيانات المرسلات عالية الجودة يمثل مشكلة صعبة في الأنظمة التي تعمل بالبطارية بسبب محدودية توفر الطاقة. تتناول هذه الرسالة هذا التحدي من خلال تحسين كفاءة الطاقة في شبكات استشعار الوسائط المتعددة اللاسلكية لبيئات المراقبة اللاسلكية ذات الموارد المحدودة. ينصب التركيز على تطوير برامج ترميز جديدة تقلل من استهلاك الطاقة مع الحفاظ على جودة عالية من الخبرة لكل من المعالجة البشرية والآلية.

تقدم الأطروحة طرقًا منخفضة التعقيد لاكتشاف مناطق الاهتمام (ROI) في إطارات الفيديو. سيعزز ذلك الدقة والمتانة من خلال الاستفادة من تقنيات اكتشاف الأجسام المتعددة. تم دمج تقنيات الكشف هذه كرميزات مسبقة في سلاسل تشفير مختلفة للمراقبة بالفيديو اللاسلكي ، مما يؤدي إلى توفير كبير في الطاقة ومعدل البت يصل إلى 98٪ مع الحفاظ على جودة خدمة مقبولة (QoS) وجودة الخبرة. تظهر العديد من الاختبارات والتجارب جدوى وفعالية الأساليب المقترحة في هذه الأطروحة. تمهد نتائج هذا البحث الطريق للبحث المستقبلي في هذا المجال.

كلمات مفتاحية: منطقة الاهتمام ، اكتشاف الأشياء ، شبكات استشعار الوسائط المتعددة اللاسلكية ، التعقيد المنخفض ، ضغط الصور ، ضغط الفيديو ، أنظمة المراقبة بالفيديو المدمجة ، جودة التجربة.

« Etude et mise en œuvre d'un encodeur vidéo basé objet pour les systèmes de vidéosurveillance sans fils embarqués »

Résumé :

Les systèmes de vidéosurveillance sans fil embarqués gagnent en popularité grâce aux progrès des systèmes Internet des objets (IoT) et des réseaux de capteurs multimédias sans fil. Ces systèmes ont de nombreuses applications, notamment le suivi et la surveillance de cibles militaires, les secours en cas de catastrophe, la surveillance de la santé biomédicale, la détection sismique, la surveillance de l'environnement et les villes intelligentes. La transmission de données multimédias à faible débit tout en maintenant des données transmises de haute qualité est toujours un challenge dans les systèmes alimentés par batterie en raison de l'énergie limitée. Cette thèse relève ce défi en optimisant l'efficacité énergétique dans les réseaux de capteurs multimédias sans fil pour les environnements de surveillance sans fil à ressources limitées. L'accent est mis sur le développement de nouveaux codeurs qui minimisent la consommation d'énergie tout en maintenant une qualité d'expérience (QoE) élevée pour le traitement humain et machine.

La thèse introduit des méthodes de faible complexité pour détecter les régions d'intérêt (ROI) dans les images vidéo. Cela améliorera la précision et la robustesse en tirant parti de plusieurs techniques de détection d'objets. Ces techniques de détection sont intégrées en tant que pré-encodeurs dans différentes chaînes d'encodage pour la vidéosurveillance sans fil, permettant d'importantes économies d'énergie et de débit jusqu'à 98% tout en préservant une qualité de service (QoS) et QoE acceptable. Plusieurs tests et expérimentations démontrent la faisabilité et l'efficacité des approches proposées dans cette thèse. Les résultats de cette recherche ouvrent la voie à de futures recherches dans ce domaine.

Mots clés : Région d'intérêt, détection d'objets, réseaux de capteurs multimédias sans fil, faible complexité, compression d'images, compression vidéo, systèmes de vidéosurveillance embarqués, qualité d'expérience.

« Study and Implementation of an Object-based Video Pre-encoder for Embedded Wireless Video Surveillance Systems »

Abstract :

Embedded wireless video surveillance systems are gaining widespread popularity due to advancements in internet of things (IoT) systems and wireless multimedia sensor networks. These systems have numerous applications, including military target tracking and surveillance, disaster relief, biomedical health monitoring, seismic sensing, environment monitoring, and smart cities. Transmitting multimedia data at a low bitrate while maintaining high-quality transmitted data is still a challenging problem in battery-powered systems due to limited energy availability. This thesis addresses this challenge by optimizing energy efficiency in wireless multimedia sensor networks for resource-constrained wireless surveillance environments. The focus is on developing novel encoders that minimize energy consumption while maintaining high Quality of Experience (QoE) for both human and machine processing.

The thesis introduces low-complexity methods for detecting regions-of-interest (ROI) in video frames. This will enhance accuracy and robustness by leveraging multiple object detection techniques. These detection techniques are integrated as pre-encoders in different encoding chains for wireless video surveillance, resulting in significant energy and bitrate savings of up to 98% while preserving acceptable quality of service (QoS) and QoE. Several tests and experiments demonstrate the feasibility and effectiveness of the proposed approaches in this thesis. The findings of this research pave the way for future research in this field.

Keywords : Region-Of-Interest, Object Detection, Wireless Multimedia Sensor Networks, Low-Complexity, Images Compression, Video Compression, Embedded Video Surveillance Systems, QoE.

This thesis work is dedicated to my beloved parents, Halima and Mohamed, who have been my unwavering support system throughout my academic journey. Their encouragement and belief in me have driven me to give my all and reach the finish line. My gratitude extends to all my brothers, my twin Houcine, my sister, and my dead brother Abdellah, who have all played a significant role in my life and have been affected by my academic pursuits in their unique ways. I cannot express enough how much I cherish and love each and every one of you. Baraka Allahu Fikom ...

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge **Allah** Almighty for granting me the health and bravery to make this contribution. And I ask Allah to guide succeed the rest of the way.

I would like to express my deepest appreciation to my advisor, **Dr. Nasreddine Kouadria** for the all-encompassing support he has provided during my doctoral thesis. His invaluable guidance and research vision have deeply inspired me and significantly improved the quality of my research work. Besides being the doctoral advisor, he also provided me with encouragement and patience throughout the duration of this project.

I would also like to extend my deepest gratitude to my co-supervisor **Dr. Saliha Harize** for her support and mentorship during all the years of the Ph.D. thesis. She acted the role of the mentor and the advisor in the difficult moments and helped me surpass the difficulties.

The successful completion of my dissertation is a testament to the good guidance and support of **Dr. Moufida Maimour**. I am deeply grateful for her unwavering dedication and commitment to providing expert advice and direction. Her tireless efforts, both as coauthor and beyond, have been instrumental in helping me reach this milestone.

I thankfully acknowledge **Professor Nouredine Doghmane** for his mentorship, assistance, and technical support during my thesis years, despite his numerous commitments.

I also wish to thank my dissertation committee president, **Professor Benouaret Mohamed**, and all the committee members, **Dr. Boumehrez Farouk**, **Dr. Boukaache Abdenour**, and **Dr. Moufida Maimour** for investing their valuable time and effort in evaluating my work. I am especially thankful to **Islem Mansri**, **Zakaria Naili**, **Abd el jalil Kebir**, **Mohamed Mekahlia**, **Oussama Hadoune** and **Housseem maatar** for being great friends and colleagues and sharing feelings with me in both good and bad situations during the thesis journey. I will forever remain grateful for their inspiration, suggestions, and confidence in me which motivated me to undertake the challenges.

I extend my special thanks to my big brother **abdelghani** and my twin **Houcine** for being my special friends during difficult moments.

Finally, my greatest thanks to my parents, and my brothers for their continuous love, care, and support. I also thank my sister for supporting my back in all aspects of my life.

List of Figures

1.1	Video surveillance systems categories and applications	9
1.2	Sensevid architecture [27]: One of the useful platforms for video coding in WMSN.	12
1.3	Results of some well-known pixel-based MOD algorithms (Highway #790)	13
1.4	A conceptual example of background modeling circuit embedded in FPGA board (redrawn from [61])	14
1.5	Wavelet-based background modeling for MOD proposed in [58].	15
1.6	ROI detection and coding approach proposed in [64].	17
1.7	WMSN model and Block decision/compression approach based on ROI proposed in [69].	17
1.8	Illustrative example of WMSN composed of visual sensors and Sink node to connect to the external network.	18
1.9	A sensor node employing ROI detection as a pre-processing step before compression	18
1.10	Existing axes of video and image coding in WMSN with some examples of approaches	22
1.11	Standard coding scheme, Compress-Then-Analyse (CTA)	23
1.12	Content-aware coding as alternative, Analyse-Then-Compress (ATC) paradigm	23

1.13	Complete chain of the approaches and paradigms involved in this thesis	24
2.1	Impact of 2-D Rank Order Filter: from Right to left: Before, After	31
2.2	Block diagram of the proposed ROI Detection.	32
2.3	Block diagram of the proposed video coding strategy.	33
2.4	Example of the effect of wrong detection on the visual results degradation and error propagation. Frame #150 of traffic2 sequence)	41
3.1	Block diagram of the proposed algorithm (BIRD)	44
3.2	FGS eliminates unnecessary activities and ROF enhances the non-zeros scores prior to thresholding	46
3.3	Impact of the combination of FGS and ROF on the ROI classification . .	46
3.4	ROC curve and the optimum threshold for <i>pedestrians</i> , <i>Highway</i> and <i>Snowfall</i> sequences	54
3.5	Number of blocks belonging to the ROI according to the threshold value	55
3.6	Per-frame energy dissipation of BIRD for <i>Highway</i> , <i>pedestrians</i> and <i>Snowfall</i>	59
4.1	Proposed ROI-based video compression scheme	64
4.2	Example shows the results of segmentation for the 3 regions, 1-Frame 85 form Hall sequence. 2-Mask of the 3rd ROI. 3-Mask of the 2nd ROI. 4-Mask of the 1st ROI (Th1=0.0224, Th2=0.0301)	66
4.3	Results of multi-QF based coding. from left to right: 1- Original Frame 2- Segmentation Results 3- Decompression results (JPEG chain with ROI1: QF=90, ROI2:QF=50, ROI3:QF=10 - PSNR=33.9308, SSIM=0.7618) 4- ROI visual quality for same bitrate(proposed left, MJPEG right).	66
4.4	PSNR value of ROI-based coding compared to MJPEG for Hall seq. . . .	67
4.5	PSNR value of ROI-based coding compared to MJPEG for traffic seq. . .	67
4.6	Mean PSNR of the whole frame and the ROI-1 for the proposed method compared to MJPEG for the same bitrate (hall seq.)	68

4.7	Mean PSNR of the whole frame and the ROI-1 for the proposed method compared to MJPEG for the same bitrate (traffic seq.)	69
4.8	SSIM results of ROI-based coding compared to MJPEG for hall seq. . . .	69
4.9	SSIM results of ROI-based coding compared to MJPEG for traffic seq. . .	70
4.10	Required bitrate for proposed strategy against standard MJPEG for traffic seq.	70
4.11	Required bitrate for proposed strategy against standard MJPEG for hall seq.	71
5.1	The scheme of the proposed strategy for a 3-level ROI-based video coding.	75
5.2	ROI construction and <i>GMR</i> mask use for extra blocks elimination. . . .	77
5.3	The proposed scheme illustration with $w_1 = 8$, $w_2 = 4$ and $w_3 = 2$	77
5.4	Affiliation of each pixel to the classified regions.	78
5.5	Impact of the previous frame selection on TPR for highway sequence . .	80
5.6	PSNR results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.	85
5.7	SSIM results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.	86
5.8	VIF results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.	87
5.9	BRISQUE scores for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight	88

5.10	Amount of data to transmit per frame for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight	90
5.11	Tiny-YOLOv3 network Architecture (redrawn from [151])	92
5.12	Performance of recognition: Number of detected objects	93
5.13	Performance of recognition: Recognition accuracy	93
5.14	Total processing energy consumption and the corresponding PSNR for sequences with different frame sizes. <i>campus</i> 352×288 , <i>highway</i> 320×240 , <i>traffic</i> 160×120	97

List of Tables

1.1	Summary of the related work on ROI-based video coding	25
2.1	Details of the used dataset	34
2.2	Used parameters and values for the simulation	34
2.3	Visual binary mask for the ROI detection.	36
2.4	PSNR, SSIM, MS-SSIM and VIF results for the used dataset	37
2.5	Ratio of data reduction using the proposed strategy	39
2.6	Mean number of blocks to be transmitted for each strategy	40
2.7	Execution Time in milliseconds for different frame size	42
3.1	Used parameters for the conducted simulations	48
3.2	Samples of ROI extraction mask results	50
3.3	Detection results of the proposed algorithm over CDnet 2014 dataset . .	51
3.4	Comparison of BIRD with classical techniques over CDnet 2014 dataset .	52
3.5	Category-wise comparison of BIRD to state-of-the-art on CDnet 2014 dataset	53
3.6	Statistics of the energy gain under threshold variation	55
3.7	ARM Cortex M3 characteristics	56
3.8	Computational budget of each step of BIRD algorithm	57

3.9	Per-frame $E_{detection}$ cost of the method compared to state-of-the-art for size (240x320)	58
4.1	Parameters and methods used for each Step	65
5.1	Used video sequences	79
5.2	Visual binary mask for the moving region.	82
5.3	Overall mean quality metrics	83
5.4	Bounding box insertion results for the used dataset.	91
5.5	Computational Cost of each step	95
5.6	Per frame energy cost (mJ) of our ROI detection	97
5.7	Per-frame energy consumption (mJ).	97

Nomenclature

Acronyms / Abbreviations

ATC	Analysis Then Compress
BAC	Balanced-Accuracy
BIRD	Block-based movIng Region Detection
BRISQUE	Non-reference
BS	Background Subtraction
CTA	Compress Then Analysis
DCT	Discrete Cosine Transform
DTT	Discrete Tchebichef Transform
ED	Edge Detection
FD	Frame Difference
FGS	Fast Global Smoother
FoV	Filed of View
FP	False Negative

FP	False Positive
GMM	Gaussian Mixture Model
GMM	Gaussian of Mixture Model
GMR	Global Moving Region
GOP	Group of Pictures
HoG	Histogram of Gradients
KDE	Kernel Density Estimation
KNN	K-Nearest Neighbors
MOD	Moving Object Detection
MoG	Mixture of Gradients
MS-SSIM	MS-Structural Similarity
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
ROF	Rank Order Filter
ROI	Region-of-Interest
SAD	Sum of Absolute Difference
SFD	Sum of Frames Differences
SSAD	Successive Sum of Absolute Difference
SSIM	Structural Similarity
TN	True Negative

TP	True Positive
TPR	True Positive Ratio
VCM	Video Coding for Machines
VIF	Visual Information Fidelity
VQA	Video Quality Assessment
VS	Visual Sensor
WVS	Wireless Visual Surveillance
WMSN	Wireless Multimedia Sensor Networks

List of Publications

Peer-reviewed Journal Publications

- [J1] Ahcen Aliouat, Nasreddine Kouadria, Moufida Maimour, Saliha Harize, and Nouredine Doghmane. "Region-of-interest based video coding strategy for rate/ energy-constrained smart surveillance systems using WMSNs." *Ad Hoc Networks* 140 (2023): 103076.
- [J2] Ahcen Aliouat, Nasreddine Kouadria, Saliha Harize, and Moufida Maimour. "An Efficient Low Complexity Region-of-Interest Detection for Video Coding in Wireless Visual Surveillance." *IEEE Access*, 11, 26793-26806.
- [J3] Ahcen Aliouat, Nasreddine Kouadria, and Doru Florin Chiper "x-DTT: A package for calculating Real and Integer Discrete Tchebichef Transform kernels based on Orthogonal Polynomials" *SoftwareX* journal (Minor revision).
- [J4] Ahcen Aliouat, Nasreddine Kouadria, Moufida Maimour, and Saliha Harize. "EVBS-CAT: Enhanced Video Background Subtraction with a Controlled Adaptive Threshold for Constrained Wireless Video-surveillance" *Journal of Real-Time Image Processing* (Under review).

Peer-reviewed Conference Publications/Proceedings

- [C1] Ahcen Aliouat, Nasreddine Kouadria, Moufida Maimour, and Saliha Harize. "Region-of-interest based video coding strategy for low bitrate surveillance systems." In *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*, pp. 1357-1362. IEEE, 2022.
- [C1] Ahcen Aliouat, Nasreddine Kouadria, Saliha Harize, and Moufida Maimour. "Multi-threshold-based frame segmentation for content-aware video coding in WMSN." In *Advances in Computing Systems and Applications: Proceedings of the 5th Conference on Computing Systems and Applications*, pp. 337-347. Cham: Springer International Publishing, 2022.

Poster

- [P1] Ahcen Aliouat, Nasreddine Kouadria, and Saliha Harize. "Low-Cost Region-of-Interest Detection for Wireless Video Sensor Nodes" In Doctoral Days of the LASA Laboratory, UBMA, June 2021.

Contents

List of Figures	i
List of Tables	v
Nomenclature	vii
List of Publications	ix
Introduction	1
1 Background and literature review	7
1.1 Video surveillance Systems	7
1.2 Resource optimization in WSN	11
1.3 Object Detection methods	13
1.4 ROI-based video coding	16
1.5 Image/video coding for wireless surveillance	18
1.6 Evaluation Metrics used in the thesis	26
2 ROI-based video coding strategy for low bitrate surveillance	29
2.1 Introduction	29
2.2 Proposed method	29
2.2.1 Edge Detection	29

2.2.2	Sum of Absolute Differences (SAD)	30
2.2.3	2-D Rank Order Filter	30
2.2.4	Fast Global Smoother	31
2.3	Qualitative Results	34
2.4	Quantitative results	38
2.4.1	Data Reduction	38
2.4.2	Data Saving over the used Dataset	38
2.4.3	Comparison with Standard methods	40
2.5	Limits in terms of visual Quality	41
2.6	Execution time	41
2.7	Conclusion	42
3	An Efficient Low Complexity ROI Detection for Video Coding in WVS	43
3.1	Introduction	43
3.2	Proposed Method	44
3.2.1	Difference Detection	44
3.2.2	Difference Enhancement	45
3.3	Results and Discussion	48
3.3.1	Parameters and experimental conditions	48
3.3.2	Performances of BIRD over the CDnet 2014	49
3.3.3	Comparison with other techniques	49
3.3.4	Metrics of Interest: Recall, specificity and BAC	51
3.3.5	The impact of thresholding on detection	52
3.3.6	Method Complexity	56
3.3.7	Energy Budget for change detection	57
3.3.8	Energy dissipation for complete compression chain	58
3.3.9	Memory requirements	61
3.4	Conclusion	61

4 Multi-Threshold-based frame segmentation for content-aware video coding in WMSN	62
4.1 Introduction	62
4.2 Proposed method	63
4.3 Results and Discussion	65
4.3.1 Image segmentation results	65
4.3.2 Compression Quality Results	66
4.3.3 Bitrate Results and Gain	68
4.4 Conclusion	71
5 ROI-based video coding strategy for rate/energy-constrained smart surveillance systems using WMSNs	73
5.1 Introduction	73
5.2 Proposed S-SAD method	74
5.2.1 ROI Detection	75
5.2.2 ROI Compression and Transmission	78
5.3 Results and Discussion	79
5.3.1 Detection Accuracy and Visual Results	79
5.3.2 Quantitative Results: Image Quality	81
5.3.3 Bitrate Gain	84
5.3.4 The Impact of Quality Degradation on Object Recognition	89
5.3.5 Computational Complexity and Energy Consumption	94
5.3.6 Conclusion	98
General conclusion and perspectives	99
Bibliography	101

Introduction

Context and problem statement

Video surveillance systems have become a pillar technology in new-generation communication systems [1]. This last fact is due to the high advantage that offers surveillance systems to ensure security, monitoring, and prevention in critical environments. The basic paradigm of surveillance systems considers a wired surveillance camera installed in a well-studied zone. However, this approach is constrained by the existence of a wired energy/connection cable to feed the surveillance system. This circumstance has been viewed as a drawback for surveillance systems because in most cases, the areas that need to be covered cannot be wired for cables and require a wireless link [2]. This last problem has endorsed researchers to develop new surveillance systems which are fully wirelessly connected and battery-equipped, motivated by the advancement of Internet of Things (IoT) systems and wireless sensor networks (WSN).

The WSNs market is getting more and more attention and growth during the last years thanks to the solutions it gives to a plurality of communications and monitoring domains. WSNs have had enormous potential for use in a wide range of contexts, including military target tracking and surveillance [3] [4], disaster relief [5] [6], biomedical health monitoring [7] [6], seismic sensing [8], and many more. Sensor modules may cover a plurality of data sensing types, either scalar sensors like humidity sensors, motion sensors, pressure sensors, or heart rate sensors. Likewise, multimedia

sensing is possible using sensor-equipped with multimedia capturing modules, which aim to process audio, images, or video data [9]. The subsection of the WSN that cover multimedia sensing (image, video, audio) is shorthand: Wireless Multimedia Sensor Networks (WMSN). The network is anticipated to include collectively covering a small area with image, audio, or video sensors [10].

WMSN has been used as a support technology for new video surveillance systems. Benefiting from its effectiveness to cover remote, essential areas where wire cabling is impractical. For instance, WMSN-based surveillance systems can be used in agriculture to monitor the fields. It is also an intriguing alternative for particular military objectives, like monitoring and spying on the battlefield. The Industrial Internet of Things (IIoT) has also utilized WMSN for specific industrial applications including monitoring components of the production line [11] [12]. WMSN-based monitoring has been deployed by a variety of scientists to monitor and track migrating birds, forests, and lakes [13]. WMSN are composed of a plurality of sensors interconnected and equipped with visual modules to capture images. They have enabled a plurality of applications and have been involved in the development of the Internet of things (IoT) and smart edge computing. As part of the applications that use WMSN as a backbone, we find wireless video surveillance systems (WVS).

Thanks to their easier installation and flexible infrastructures, WVS systems now act as pillars of the new smart cities paradigm [14]. It assists in traffic monitoring [15], parking management and public safety protection in campuses, office buildings, or residential areas [16]. Since each sensor node in the WMSN is generally equipped with a limited source of energy (batteries), they have to keep a high presence rate to ensure a good Quality of Experience (QoE). Keeping a long-life feature to the sensor node needs to elaborate and develop low complexity methods to be embedded in different steps of the sensing from capture to transmission [17].

Additionally, reducing energy consumption is a challenging task since there is always a need for optimal design, accurate tasks and good Quality of service (QoS) while

developing such methods [18]. The effective transmission of multimedia data, particularly video and images, over constrained bandwidth and energy resources is one of the major issues faced by WMSNs. Compression techniques and transmission rate control are currently used as solutions to this problem, although they frequently lead to trade-offs between energy usage and image quality.

The limitations of surveillance systems that use WMSN must be made very explicit in order to effectively demonstrate the issue. Each node in the network has a limited energy source because they are all wirelessly connected to one another and to other networks. On the other hand, given multimedia data is so massive and demands a lot of processing power and transmission payload, wireless nodes that capture multimedia data have significant battery drain. In order to assure long-life services, it is necessary to create new and effective techniques for lowering energy consumption and data rate while keeping an acceptable level of image/video quality in WMSNs. This goal must be defined and adapted to the characteristics of the WMSN.

The early-stated problem could be transformed into many research questions in order to clarify the proposed contributions of this thesis. The thesis tries to respond to the following questions:

- How can region-of-interest (ROI) detection methods be used to reduce energy consumption in the compression and transmission steps for video/image coding in WMSNs?
- What are the most accurate and low-cost ROI detection methods that can be used in a wireless surveillance environment?
- What are the advantages and limits of such algorithms in the WMSNs Context?
- How does the accuracy of the used ROI detection method affect the effectiveness of human-based and machine-based monitoring systems at the destination?

Objectives

We aim in this thesis to focus on advancing state-of-the-art by developing accurate and efficient techniques to serve as pre-encoders in wireless sensor nodes. Those techniques will help reduce significantly the energy and bitrate needed to transmit multimedia data while saving high-quality transmitted data. The specific objectives of this thesis are to:

- Investigate the use of ROI-based video and image coding techniques in wireless surveillance environments.
- Develop and evaluate new ROI detection methods that can be used to reduce energy consumption and data rate while maintaining good QoS in terms of image/video quality.
- To evaluate the importance of accurate ROI detection methods in allowing for both human-based and machine-based monitoring at the destination, in order to improve the overall effectiveness and efficiency of wireless surveillance systems.
- Compare the performance of the proposed methods to existing state-of-the-art techniques.

Figure (1.13) illustrates a global view of this thesis's contributions to a WVS system.

Thesis Outline

The present thesis is organized in the following manner:

- **Introduction:** This chapter outlines the problem addressed by the thesis, which is the optimization of wireless surveillance systems. The significance of exploring the challenges of video coding in resource-constrained wireless surveillance systems, the current research gap in the literature, and the state of the art are

emphasized. This chapter also highlights the proposed approaches taken in this thesis to achieve the desired objectives.

- **Review of Literature:** In this chapter, we present a thorough examination of the background and research efforts made in video surveillance systems. We also present a deep examination of the works that aimed at designing accurate and cost-effective video coding strategies that are optimized for sensor node constraints. We provide a comprehensive overview of the contributions made in the processing step and transmission step of sensor nodes. Furthermore, we delve deeper into the literature that employs moving object detection and region of interest detection as pre-encoders in WVS. We also focus on low-cost ROI detection methods that help minimize resource consumption in sensor nodes.
- **Second chapter:** This one proposes a binary classification approach for frame blocks, where the blocks are classified as either ROI or non-ROI. The proposed ROI detection strategy uses edge features and difference enhancement, with continuous frame updates. We evaluate the efficiency of this method by comparing it with a standard coding approach and demonstrate significant energy savings through energy/rate efficiency modeling and analysis. However, we also acknowledge the limitations of the method, particularly with regard to the error propagation problem during reconstruction.
- **Third chapter:** In this chapter, we make a proposal for improving wireless video surveillance by classifying frame blocks into binary classes (ROI/non-ROI) using a simple difference detection method and a combination of enhancing filters. We demonstrate the effectiveness of this approach by comparing it to a range of state-of-the-art methods using a large benchmarking dataset. Our method proves its efficiency through energy/rate efficiency modeling and analysis as input to the video encoder.

- **Fourth chapter:** We propose a content-aware multi-class frame block classification method based on edge feature detection and enhancement. This method classifies the frame into three regions based on importance, prioritizing the moving object by allocating the highest bitrate to it. In this chapter, we demonstrate the efficiency of the automatic thresholding method used and show a 50% reduction in bitrate through consideration of the important region and its adaptability to WMSNs.
- **Fifth chapter:** In this chapter, we propose a novel video coding method optimized for both the machine inference model and the human visual system in WMSN. Our approach utilizes a low-cost and accurate ROI detection and classification technique to divide the frame blocks into four classes based on their importance. The video encoder then follows the ROI recommendations to decide whether to code or drop a block and adjust the coding quality accordingly. The results show that our method can achieve a 96% reduction in bitrate, a 98% reduction in energy consumption, and a 22% improvement in the deep learning-based inference model at the destination, despite a quality reduction of 1 to 4dB.
- **Conclusion:** This last chapter concludes the thesis and discusses future perspectives.

Background and literature review

The literature has widely discussed the resource optimization techniques used in WSN from different points of view. Also, the feasibility of WMSN as a backbone of WVS systems has been discussed and developed, as will be shown in the next section. From another part, ROI-based video coding has also been promoted as a promising solution for developing accurate and energy-efficient video and image coding embedded systems to leverage the bandwidth and energy constraints of WMSN. Moreover, since the ROI is defined by the moving part in the frame, it has been defined and developed based on MOD algorithms, as to be shown in the literature review.

1.1 Video surveillance Systems

Video surveillance is an essential mechanism for monitoring and observing activities in various areas, including public spaces and sensitive locations such as private enterprises and buildings. The underlying technical processes involve capturing real-time images using deployed cameras and transmitting the acquired data through a suitable transmission channel. The received signal is then reconstructed and displayed at a centralized location for analysis and assessment.

With advancements in video technology, various video surveillance systems have been developed, each possessing its unique features and capabilities. These systems

encompass traditional analog CCTV (Closed-Circuit Television) surveillance systems, digital networked video recorder (DVR) systems, networked IP camera systems, cloud-based systems, and decentralized mobile systems. Despite the diversity in these systems, the integration between them can pose challenges due to the use of different equipment and models. Various methods have been proposed to connect the cameras to the base station and facilitate data collection to mitigate this issue. These connections can be classified into three categories: analog, digital, and network connections between the central station and the cameras [19] [20] [1]. Figure (1.1) summarizes the applications and categories of video surveillance systems.

Analog surveillance system It is the basic model for image transmission, exchange, and recording in analog signal processing. The analog system uses coaxial cable and long-distance transceiver optical fiber [21]. The analog system shows advantages and drawbacks. Its main drawback is the limited transmission distance, while it represents advantages in image restoration capabilities.

Digital surveillance system In digital surveillance systems, MJPEG, MPEG-4 and H264 video coding standards are used at the edge to ensure low bitrate and bandwidth requirements, making it more efficient than the traditional analog system. In this technology, the cameras are generally connected through an IP network and transmission is possible using network protocols. Furthermore, the pillar of the continued advancements for digital surveillance systems is the parallel advancement and innovations in signal processing technologies and techniques in capturing, coding, transmission and representation [22].

Network surveillance system Networked cameras are a type of digital video camera that collaborate in a specific network to cover, transmit, and exchange video data. The network comprises many camera terminals and a piece of master equipment with

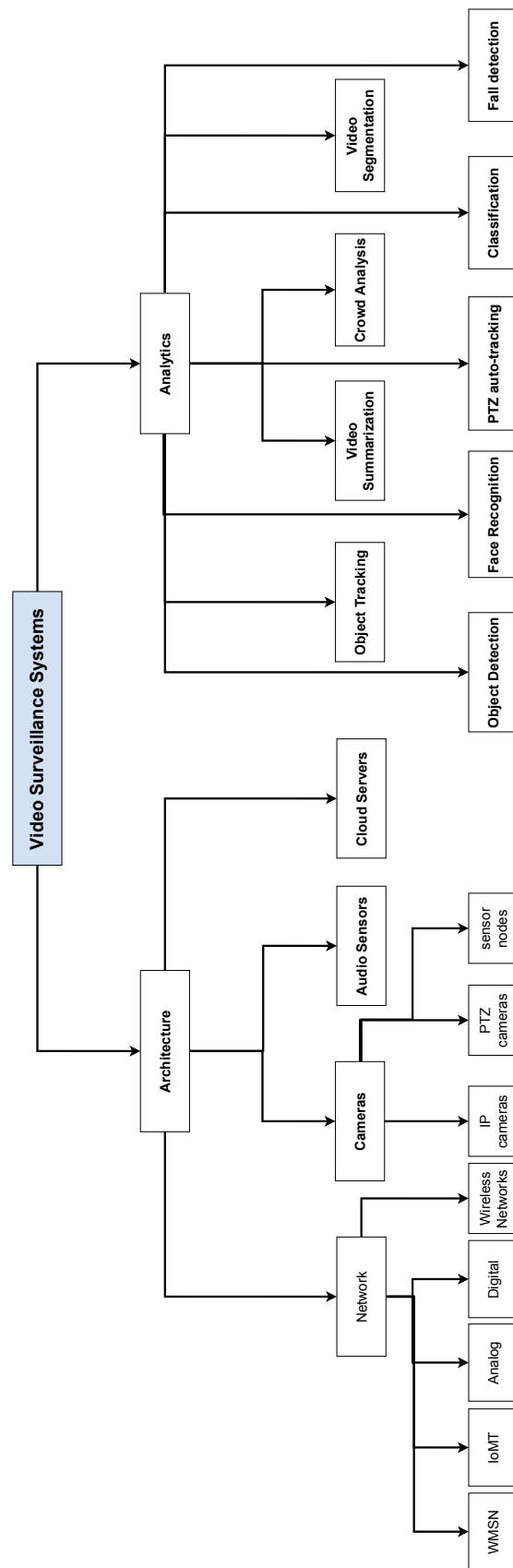


Figure 1.1: Video surveillance systems categories and applications

sufficient resources to connect the camera network to external networks [23].

Wireless Sensor Networks as surveillance system A surveillance system based on WMSN is a cutting-edge decentralized system that harnesses the power of multiple camera nodes to gather and analyze visual data. These nodes work together to extract valuable information about the captured scene and provide real-time insights to the user. One of the key features of a Wireless Video Sensor Network is its wireless communication capability, which enables the camera nodes to interact and exchange data with other nodes without the need for physical connections. This last advantage can simplify the deployment process and makes the system highly flexible and scalable. Additionally, the nodes in a WMSN typically employ advanced image processing techniques to identify and track objects, recognize patterns, and provide detailed analysis of the captured scene. This enables the system to deliver sophisticated and accurate information to the end user, making it an indispensable tool in a plurality of applications [24] [25]. Despite its promising advantages, surveillance systems that use WMSN are facing serious challenges, we cite:

- **Limited Bandwidth:** WMSNs have limited bandwidth, which can limit the amount of data that can be transmitted and received.
- **Power Constraints:** WMSNs typically rely on batteries to power the camera nodes, which can limit their lifespan and require frequent battery replacements.
- **Security Concerns:** WMSNs are vulnerable to security threats such as hacking, eavesdropping, and data tampering, which can compromise the confidentiality and integrity of the transmitted data.
- **Complex Deployment:** Setting up and maintaining a WMSN can be complex in specific zones like the wild and the lacks.

- **Cost:** Implementing a WMSN can be expensive, as it requires deploying numerous camera nodes and installing wireless infrastructure.
- **Limited Range:** Wireless communication has a limited range, which can limit the coverage area of the Video Sensor Network and make it difficult to deploy in large or complex environments.

1.2 Resource optimization in WSN

The literature has extensively discussed and analyzed the design of energy-efficient Wireless Sensor Networks [26] [27] [28] [29]. 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. Multiple contributions have been made in this context. The Medium Access Protocols (MAC) design has been the subject of optimization to meet the requirements of both energy efficiency, delay reduction and QoS insurance [30] [31]. The design of such protocols under these constraints is a complicated task since they require a continuous data stream. For example, in [32], the Saxena protocol is proposed to meet the QoS requirements for video streaming in WMSNs. Diff-MAC [33] is a QoS-oriented MAC protocol for WMSNs with heterogeneous traffic classes. It employs a service differentiation mechanism that fragments long video frames into smaller video packets and transmits them in bursts. The contention window size and duty cycle of the node are adjusted based on the traffic class. The protocol delivers data fairly and quickly and adjusts to changing network conditions, but incurs overhead from its differentiation mechanisms and network monitoring. It also lacks sleep-listen synchronization between neighboring nodes.

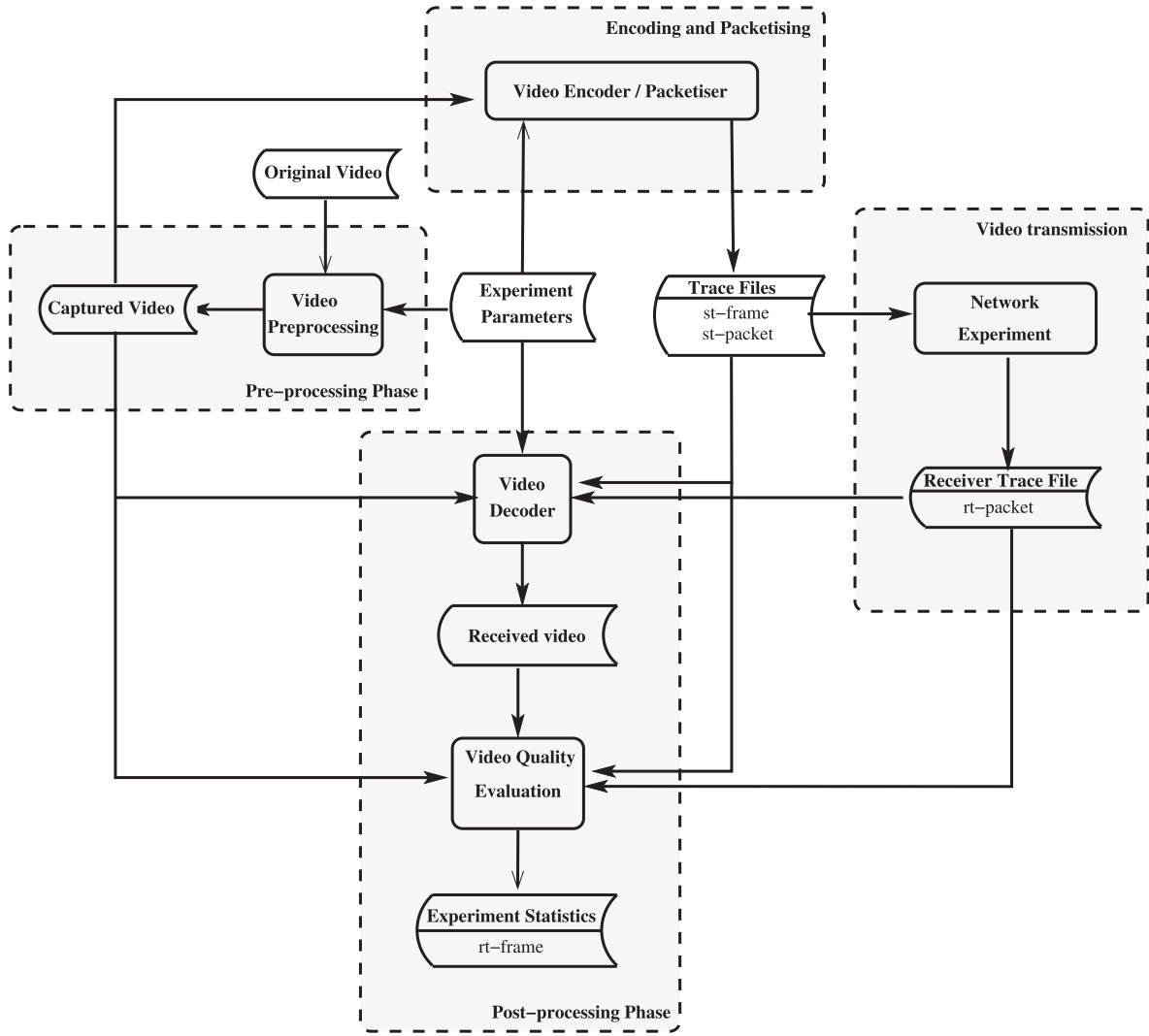


Figure 1.2: Sensevid architecture [27]: One of the useful platforms for video coding in WMSN.

Other platforms have been created specifically for low-bitrate video coding in WMSN, such as SeneVid [34] developed by Maimour. As shown in Figure (1.2), this platform is designed to meet the needs of WMSN, with different modules for video coding and a low-cost mechanism. It offers fast transforms for video compression, both exact and fast-pruned versions. It can provide real-world scenarios using the widely used WSN test bed, IoT-LAB [35]. Other platforms for WSMN include Cyclops [36] and XYZ-ALOHA [37].

Another ax of resource optimization is the utilization of low-cost embedded transformations for data compression. Literature has proposed fast and energy-efficient

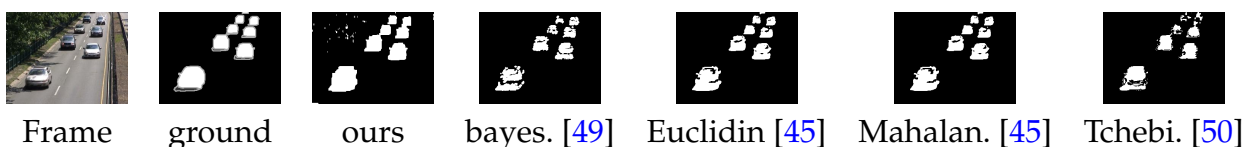
transformations to minimize energy consumption during the transform stage of compression while retaining a minimum level of quality. Efficient techniques include DCT [38] [39] [40] and its variants. DTT integer version and its variants [41] [42], and numerous other transformation methods tailored for WSN.

1.3 Object Detection methods

There is a wealth of movement detection techniques, as surveyed in numerous studies such as [43], [44], [45], and [46], which examine the latest approaches for detecting motion in video sequences. These methodologies encompass basic techniques, statistical methods, fuzzy logic, neural networks, wavelet-based background modeling, background clustering, and background estimation-based MOD detection [47]. Other researchers have focused on enhancing specific steps in the MOD process. For instance, Bouwmans et al. proposed a taxonomy for background initialization in [48], which classifies the MOD area into various categories based on methodology, recursiveness, and selectiveness. The authors also emphasized the significance of background subtraction (BS) algorithms in numerous applications, such as video compression, video surveillance, video segmentation, and video inpainting.

Recent studies have explored BS for movement detection. [51], for instance, presents a BS approach in which the background model is constructed using the Gaussian Mixture Model (GMM) algorithm. For background estimation, the authors utilized a Gaussian mixture to model the pixel intensity values. Another background modeling tech-

Figure 1.3: Results of some well-known pixel-based MOD algorithms (Highway #790)



nique was introduced in [52] and [53], where Zhao et al. modeled the background using a Type-2 Fuzzy Gaussian Mixture Model, which accounts for the uncertainty related to information or noise and addresses the limitations of the GMM model, particularly for infrared videos. Figure 1.3 illustrate results of some well-known pixel-based MOD algorithms.

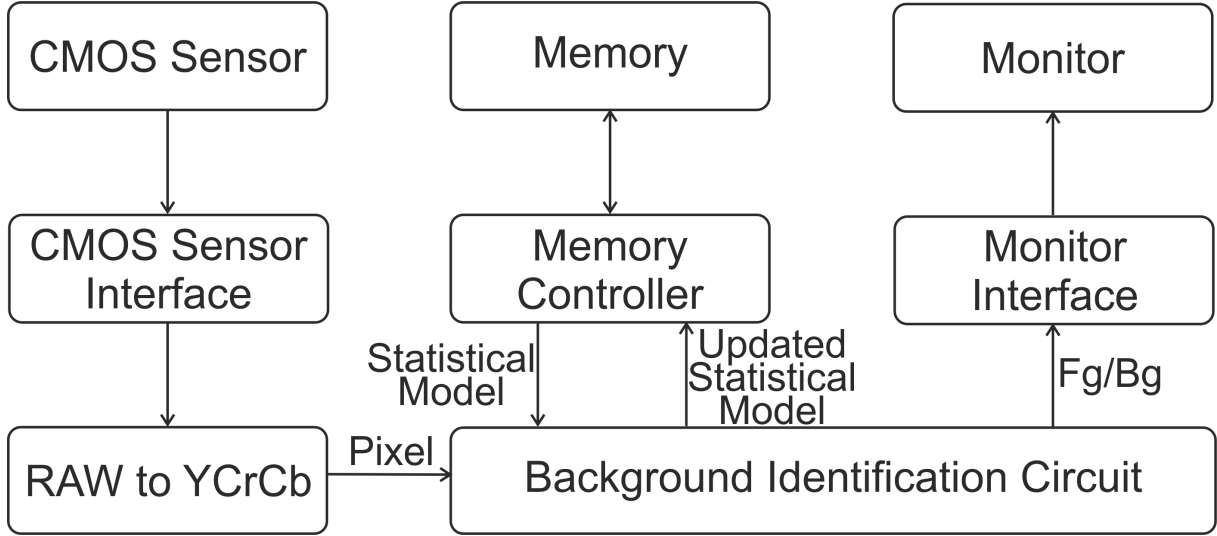


Figure 1.4: A conceptual example of background modeling circuit embedded in FPGA board (redrawn from [61])

Recent advances in Convolutional Neural Networks (CNN) and deep learning (DL) have also been applied to the MOD problem. [54], [55], [56], and [57] are examples of such techniques. In [54], the author utilized a CNN network to extract spatial information from different neighborhoods of pixels. Another solution was proposed in [55], where the authors employed DL in an unsupervised manner, using a greedy layer-wise pre-training strategy and a conjugate gradient-based backpropagation algorithm for network fine-tuning. The field of background modeling has seen a surge of interest in advanced AI techniques, as demonstrated by Babaee et al.'s deep learning approach for background modeling based on Convolutional Neural Networks [56]. Despite the high detection accuracy achieved by these DL algorithms, they are still far from being suitable for real-time, low-power embedded applications due to their high computational complexity and energy consumption. In [54], the authors attempted to use embed-

ded deep learning for motion detection, utilizing a Neural Response Mixture (NeRM) model to extract the features of the background and detect motion. Although the approach yields good detection results, it still falls short of real-world requirements, with approximately 2 fps on a 352×240 video frame when implemented on Axis Q7436 (ARTPEC-5 chipset) encoder.

Other techniques employ wavelet-based background modeling for moving object detection, as discussed in [58], [59] and [44]. For example, in [58], the authors used the wavelet domain to detect moving objects as illustrated in Figure (1.5). Similarly, [59] proposed a region of interest-based technique that models the background by characterizing regions using a mixture of multiple Gaussian modes and wavelet coefficients. The suitability of MOD approaches for reducing the storage and energy consumption of surveillance systems has been demonstrated, as in [60], where Hamed et al. proposed a power-efficient, real-time embedded realization of adaptive vision algorithms. Hung-Yu et al. [61] also proposed a hardware-oriented algorithm for object detection based on BS, implemented on an FPGA platform as illustrated in Figure (1.4). The method achieved 56 fps for the whole system and 348 fps for the BS module. In [62], authors proposed low-cost vehicle detection techniques for video-surveillance systems.

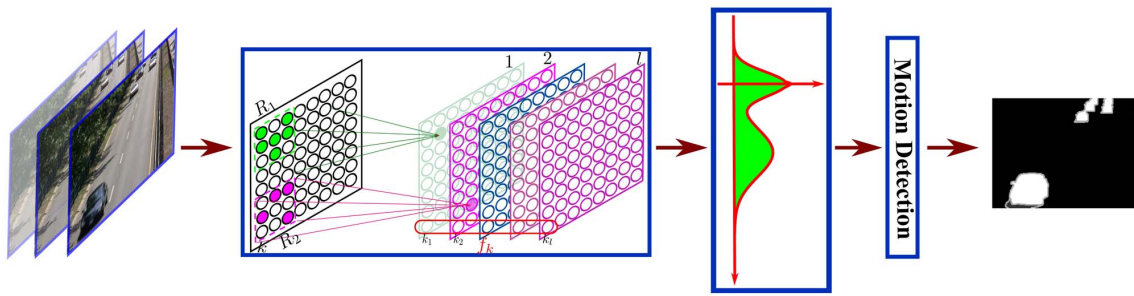


Figure 1.5: Wavelet-based background modeling for MOD proposed in [58].

Those advancements in object detection models for video coding in low-power systems have produced promising results. The integration of the MOD method, as demonstrated in works such as [63], [64] [65] and [66] has successfully reduced band-

width costs and energy consumption. These strategies have proven effective in achieving optimal results for adaptive video coding, compression, and video transmission in low-power environments.

Using MOD in WVS has experienced significant growth. A variety of techniques have emerged, ranging from simple approaches to advanced artificial intelligence methods like deep learning. Numerous surveys have explored state-of-the-art movement detection techniques, with a focus on enhancing key steps in the process and minimizing the storage and energy consumption of surveillance systems. The current research efforts are geared towards finding the optimal balance between high detection accuracy and practical considerations, such as low energy consumption and high frame rates. There is still a lot of room for improvement in this area and great potential for these techniques to be applied in the context of video coding in wireless surveillance environments using WMSN.

1.4 ROI-based video coding

ROI-based video coding strategically optimizes the compression and transmission of the frame based on moving ROI [67] [68]. ROI-based video coding employs two key approaches to achieve its objective. Firstly, it selectively compresses and transmits only the moving ROI blocks, discarding non-ROI blocks to conserve energy and bitrate, as has been proposed in [64] [63] [69].

Secondly, other approaches classify the frame blocks into multiple priority classes and allocate resources accordingly. This approach assigns higher Quality Factors (QF) to high-priority blocks to preserve image quality, lower QF to low-priority blocks and discards non-active blocks. Figure (1.6) shows the ROI detection and compression scheme proposed in [64] which is based on activity detection using SAD operation between a background model and the current frame. Then, the moving blocks are compressed using an embedded fast integer DTT algorithm. Furthermore, another

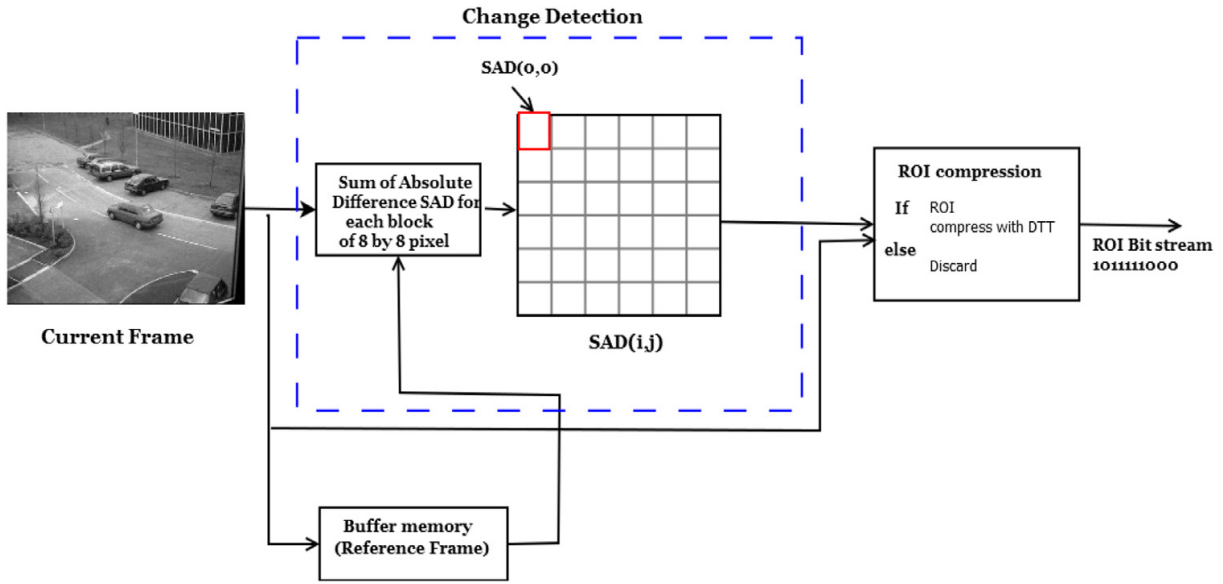


Figure 1.6: ROI detection and coding approach proposed in [64].

ROI-based video coding approach in WMSN is shown in Figure (1.7) which is proposed in [69]. The method considers both a WMSN design and ROI detection and compression using 2D-DWT by dividing the frame into 4 non-overlapping blocks.

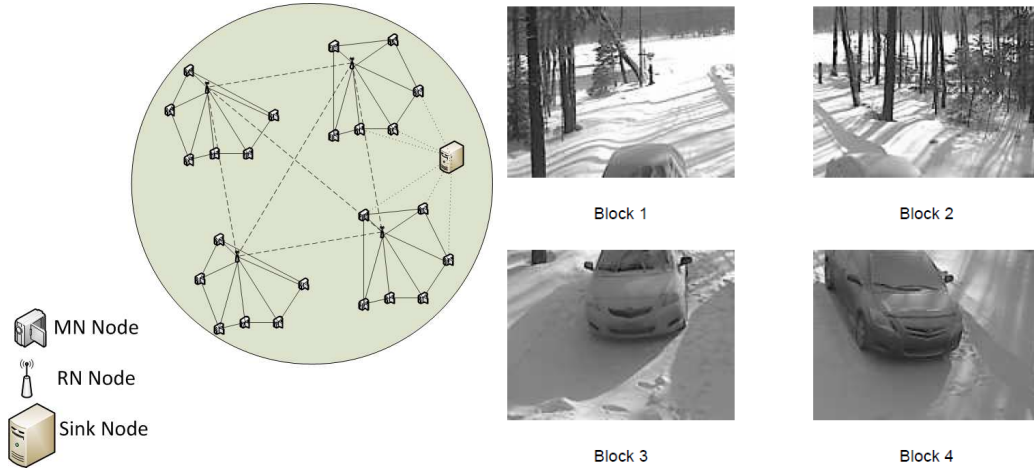


Figure 1.7: WMSN model and Block decision/compression approach based on ROI proposed in [69].

These techniques demonstrate significant improvement in energy and bandwidth efficiency over conventional techniques such as Motion JPEG (MJPEG). It has also demonstrated outstanding results for advanced codecs used in video surveillance [70], bitrate control [71], storage enhancement, [72], video-based tracking [73] and packet

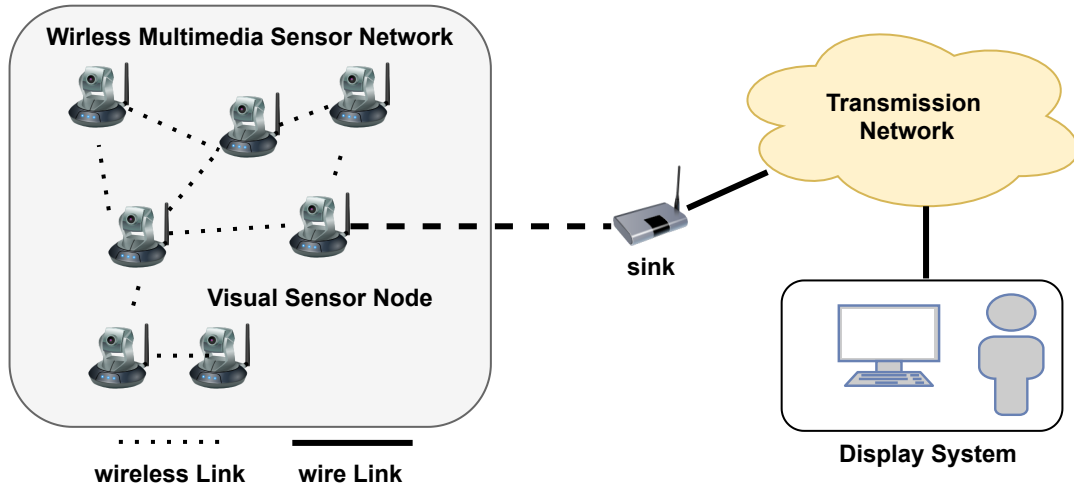


Figure 1.8: Illustrative example of WMSN composed of visual sensors and Sink node to connect to the external network.

delivery and scheduling in the wireless networks [74]. Figure (1.8) illustrates a WMSN example, while Figure (1.9) shows how the ROI detection approaches are added to the processing step on the wireless visual sensor.

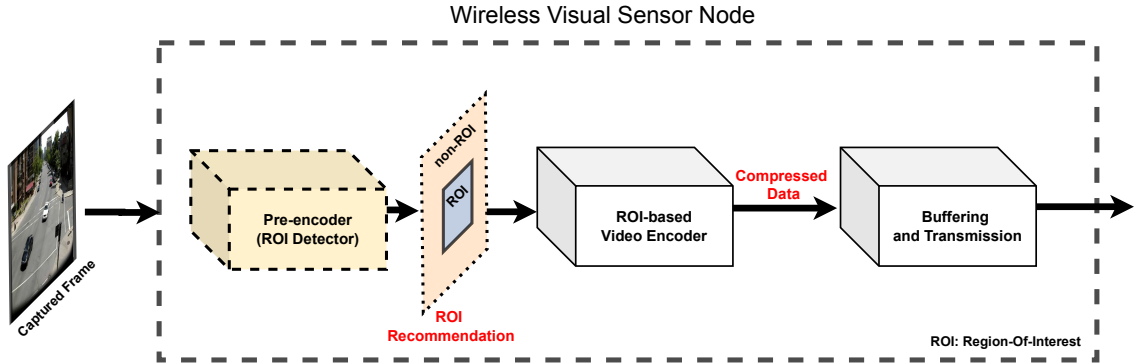


Figure 1.9: A sensor node employing ROI detection as a pre-processing step before compression

1.5 Image/video coding for wireless surveillance

The development of efficient applications for wireless surveillance requires reducing energy consumption and bitrate while keeping a high quality of transmitted information. This tradeoff is still a challenge, and there is further room for improvement in this

regard [75]. The main problem is the increased computational cost of the current video coding standards despite their high bitrate reduction and their high quality. Consequently, efforts have been made to implement the recent video coding standards in WMSN [76].

The computational budget makes the recent standards unsuitable for resource constrained sensor nodes (limited energy, limited memory and a low bitrate) [77].

In [78], a survey on energy-efficient compression and communication techniques for multimedia in resource-constrained systems has been proposed. The authors focus on the compression and processing part by comparing three main techniques of image compression in resource-constrained systems, namely: JPEG, SPIHT, and JPEG2000. By modeling energy consumption as an optimization problem of the sum of the operations made during the processing (named layers) to reduce the distortion, several optimization solutions to the problem have been presented. The first attempt was using dynamic programming [79], a faster linear-complexity algorithm to maximize the expected error-free source rate presented in [80]. Authors in [81] [82] reduced the complexity into linear complexity using a local search algorithm. From another hand, other optimization algorithms have been presented as shown in [83] using discrete ergodic search, or theoretical solutions as shown in [84].

It has been shown that recent video coding standards, referred to as predictive video coding (PVC), are not suitable as video encoders in wireless surveillance cameras [78]. This is due to the architecture of the H264 and HEVC standards that makes the central encoder leverage huge complexity compared to the decoder, making it well-suited to downstream applications where the decoder is for instance a mobile terminal.

The Distributed Video Coding (DVC) paradigm presents a new approach to PVC standards in WVS systems. Based on information theory, DVC assumes the existence of two statistically correlated, independently and identically distributed (i.i.d) video sequences, X and Y , encoded by two separate encoders aware of each other. The decoder holds complete information about the encoders. DVC frames the cod-

ing process as an optimization problem, seeking to minimize the bitrate of the video sources while ensuring sufficient accuracy in the joint reconstruction of both video sequences at the decoder side [85]. To this end, a plurality of contributions shows the promising results of DVC as a solution to constrained wireless video surveillance systems [86] [87] [88] [89] [90].

Against this background, many recent approaches have worked on optimizing the wireless sensor node resources using ROI-based video coding approaches. These approaches aim to give high priority (higher memory space, bitrate allocation, quality and network priority) to an ROI or many ROI in the video frame while decreasing the importance of non-ROI zones [91] [92] [93] [71]. Figure (1.10) shows a brief taxonomy of the existing axes of research in video and image coding in WMSN with some examples of the proposed approach under each ax.

The first step in any ROI-based video coding technique is to detect the ROI. The ROI is typically defined by the moving region or objects. Several studies have been proposed to detect the moving region in the frame based on standard and well-known moving object detection (MOD) techniques [94]. For example, using background subtraction, the moving object can be isolated by modeling the background using well-known techniques such as GMM, Histogram of Gradient (HoG) [95], codebook [96] and ViBe [97]. The mentioned techniques perform well in MOD tasks but suffer from costly computation, 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 background subtraction (BS) [98] [99]. FD has been used for MOD and has presented advantages in low complexity, low memory and speed. But it suffers from low accurate results when dealing with noisy backgrounds [63]. Edge Detection (ED) has also come up with a solution to enhance the efficiency of the MOD algorithms; but, it could suffer from high computational costs in the used edge detection operators. Hence, A low-cost ED operator is required [66] [63].

To enhance the classical low-cost methods for MOD in resource-constrained envi-

ronments, several studies have been suggested. While the ROI includes the moving object, it has been the subject of many contributions. For example, the method in [69] aims to divide the frame into four blocks before performing coding and transmission of the blocks that contain the ROI. The proposed strategy presented reasonable energy consumption for WMSN with a low bitrate. Another approach has been proposed in [64] where the idea is to enhance the BS algorithm using the sum of absolute differences (SAD). The method aims to detect, code and transmit only the region of the frame which contains a high activity. In [63] a mixture of FD, ED and summation-based ROI detection is suggested for high ROI quality and to save channel bitrate.

The early presented methods provide good results in terms of bitrate reduction and tight energy consumption. However, some of them may suffer from low precision in the MOD part leading to a notable reduction of the image quality at the receiver. The early mentioned methods still need improvements in the detection part to guarantee a high-quality ROI at the reception.

In a WMSN-based surveillance system, the communicated video or image could be the subject of advanced tasks [100] [101]. Usually, the receiver of the sensed video is a human-based system. That presumes a base station with a human operator that monitors the received video scene. In this case, ROI-based video coding is supposed to be a useful solution to enable a useful video analysis.

Nonetheless, new approaches propose intelligent automated video monitoring in a new paradigm named Video Coding for Machine (VCM) [102] [103] [104]. Which, it is supposed that the received frames are processed by a machine-aided system. A machine-based smart surveillance system uses advanced artificial intelligence and deep learning to quickly and efficiently decide about the monitoring process and to aid in accurate decision-making [101]. VCM has been used in many areas, especially in surveillance systems. It comes with potential benefits [105]; we cite:

- Low bit-rate: considering only the transmission of pertinent features.

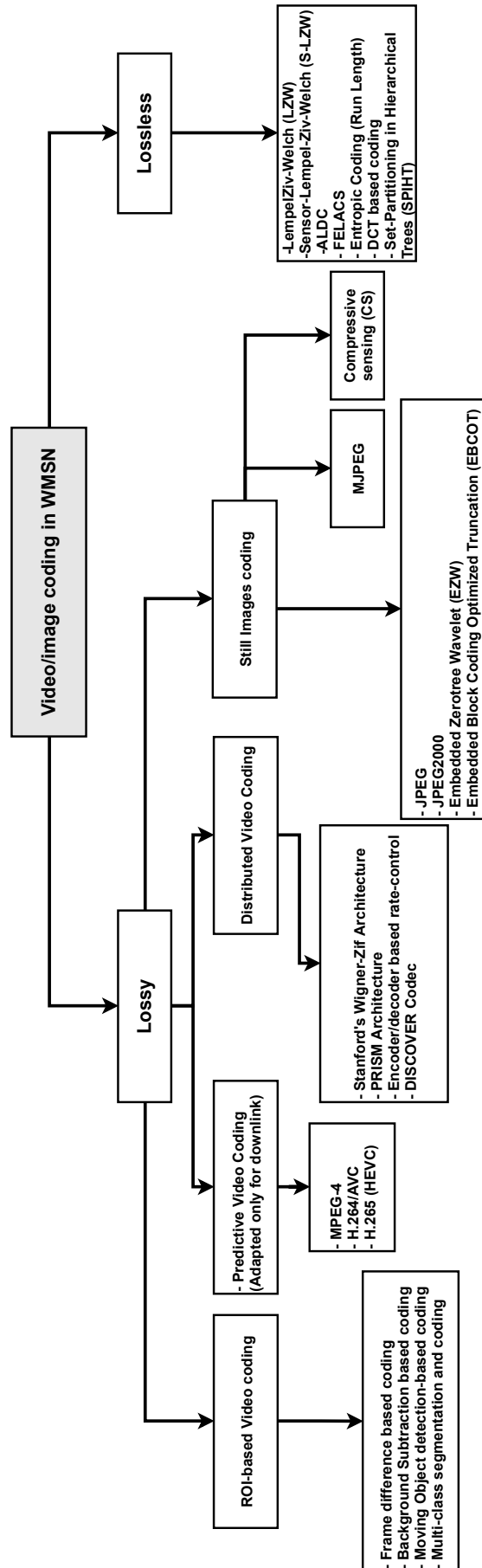


Figure 1.10: Existing axes of video and image coding in WMSN with some examples of approaches

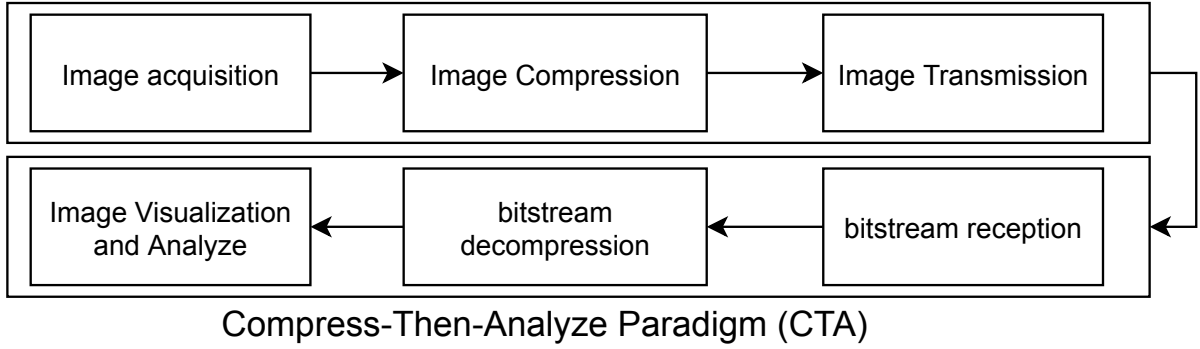


Figure 1.11: Standard coding scheme, Compress-Then-Analyze (CTA)

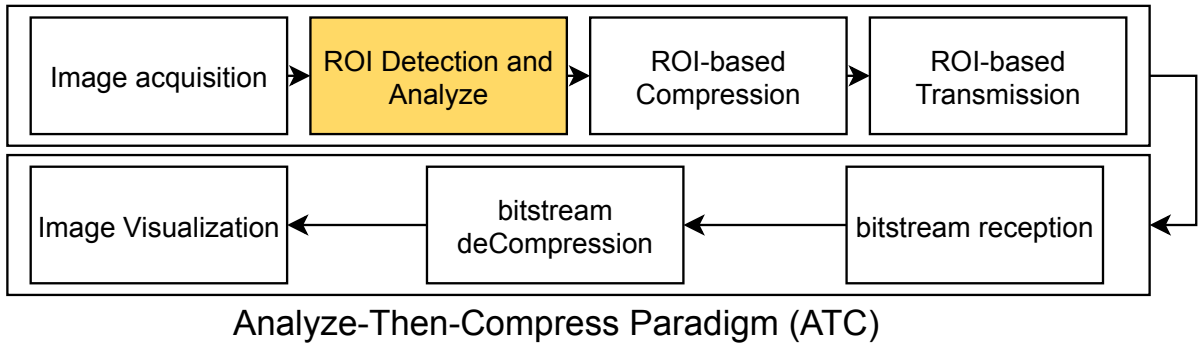


Figure 1.12: Content-aware coding as alternative, Analyse-Then-Compress (ATC) paradigm

- High precision: at the destination, the features are tuned to the need.
- High fidelity: since the data is reduced, it will be transmitted with more error resiliency.
- Balancing computation: A more balanced computation in the sensor node is achieved allowing a better lifetime.
- Privacy protection: A higher protection is expected since only some features are coded and transmitted.

The impact of lossy compression on the performance of surveillance in a VCM context has been studied. For example, in [106], authors have shown that the accuracy of the object recognition deep learning models could be affected by the low quality of the image at the reception for autonomous driven cars if the quality is lower than a certain threshold. For that reason, the guarantee of the ROI quality is justified and supposed to

enhance the efficiency of the machine-based systems at the reception. Either if the application is a machine-targeted surveillance system or a human-targeted surveillance system, the end-user interest is mostly in the information quality of the ROI. Figure (1.12) shows the application of CTA paradigm as standard coding approach, compared to ATC paradigms in Figure (1.11) as a new approach.

Recently, there has been a concerted effort to develop ROI-based video coding strategies for wireless surveillance systems that account for both human-based and machine-based monitoring at the destination, as evidenced by the works presented in [107] and [108]. The initial foray into this field was made by [67], where an ROI-based video coding surveillance system was proposed, incorporating a thorough evaluation of the efficiency of ROI-based coding for machine-based monitoring. The authors of the proposed work showcased the effectiveness of their ROI-based coding approach in a wireless surveillance system augmented by object recognition at the destination. Table 1.1 and summarize some related work on ROI-based video coding.

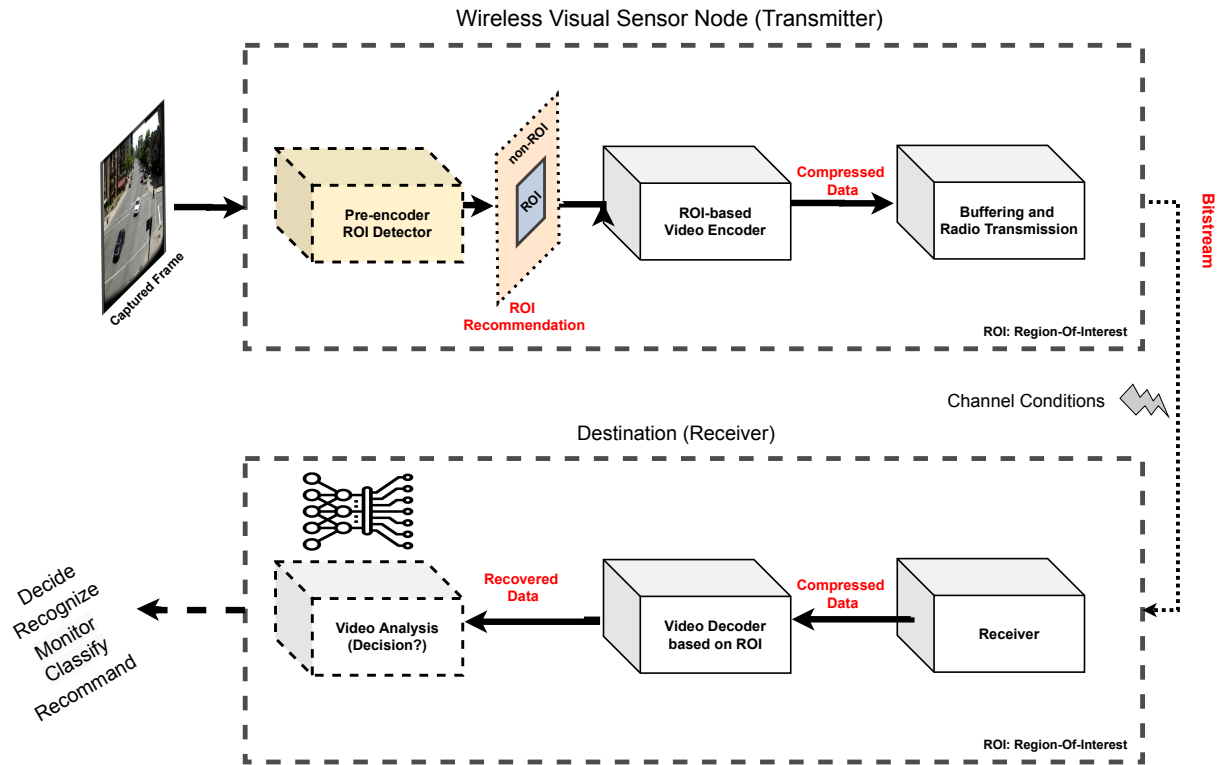


Figure 1.13: Complete chain of the approaches and paradigms involved in this thesis

Table 1.1: Summary of the related work on ROI-based video coding

Algorithm	Methodology	Highlights	Limitations
Kouadria et al. (2019) [64]	- 8×8 SAD - thresholding to extract ROI mask. - DTT transform for compression	- low complexity - fast image compression algorithm - dedicated to WMSN context	- less accurate - few datasets - few evaluation metrics
Rehman et al. (2016) [69]	- divide the frame into 4 blocks - select ROI from sub-blocks - background modeling - compression using DWT	- moderate accurate detection - simple and efficient algorithm - dedicated to WMSN context	- limited datasets - high bitrate - high complexity for WMSN node
Aliouat et al. (2022) [66]	- edge detection using Canny filter - 8×8 SAD of the edge map - automatic multi-threshold selection - multi-Otsu thresholding - compression priority to the ROI	- automatic thresholding - accurate detection - content-aware coding - allocate more resources to the ROI - dedicated to WMSN context	- high complexity - limited dataset - high bitrate (50% reduction) - no energy consumption model - few evaluation metrics
Aliouat et al. (2022) [65]	- edge detection using the Sobel filter - 4×4 SAD of the edge map - 2-D Rank order map filtering - fixed threshold - background update each GOP	- good accuracy on the used dataset - efficient in different weather cond. - high bitrate and processing reduction - dedicated to WMSN context	- high complexity for WMSN context - limited dataset - no energy consumption model - few evaluation metrics
Ko. et al. (2018) [67]	- edge detection using the Sobel filter - 8×8 SAD - bitrate control using PID-controller - optimal enhancement algorithm - prototyping on 130nm sensor node. - FPGA implementation	- accurate detection - optimal circuit design - high processing and bitrate reduction - dedicated to WMSN context	- limited dataset (2 sequences) - no comparison to the state of the art - few evaluation metrics
Ko. et al. (2015) [63]	- edge detection using the Sobel filter - perform Frame difference - 8×8 SAD - rate control (channel cond. -BER-) - thresholding using PID controller	- optimal circuit design - high processing and bitrate reduction - content and energy-aware - dedicated to WMSN context	- limited dataset (4 sequences) - no comparison to the state-of-the-art - few evaluation metrics - detection accuracy not reported
Aliouat et al. (2023) [109]	- a novel (S-SAD) introduced - multi-classes coding 2 based on ROI. - assessed for Human and Machine based monitoring	- accurate detection - energy model provided - high bitrate and processing saving - content-awareness - resources/quality tradeoff achieved - dedicated to WMSN context	- no detection accuracy comparison - medium dataset - fixed threshold
Sengar et al. (2020) [110]	- MOD detection using Optical flow - Ostu for thresholding - particle swarm optimization (PSO) for redundancy exploring	- deals with moving cameras - good efficiency compared with the state of the art - good rate-distortion performance	- limited dataset (4 sequences) - no energy consumption model - not dedicated to WMSN context - few evaluation metrics
Aliouat et al. (2023)(BIRD)	- 8×8 SFD - 1-D ROF on the activity map - FGS filter on the activity map - a pre-encoder for video coding.	- low complexity - a high detection accuracy - energy modeling (ARM Cortex M3) - large dataset (51 sequences) - dedicated to WMSN context	- tested only for fixed camera - fixed threshold

1.6 Evaluation Metrics used in the thesis

Evaluation of ROI detection Multiple metrics are used for the assessment of the ROI detection methods proposed in the thesis. Seven of them are calculated using the confusion matrix that contains the classification characteristics in terms of quality and quantity. We define and express in what next the significance of each metric as below:

TP: True positives, the number of pixels correctly labeled as belonging to the moving object.

FP: False positives, the number of pixels incorrectly labeled as belonging to the moving object.

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

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

Seven measures are substituted for the preceding four in order to more accurately assess the classification results. The metrics are given as equations (1.1 to 1.8).

Recall:

$$Re = \frac{TP}{TP + FN} \quad (1.1)$$

Specificity:

$$Sp = \frac{TN}{TN + FP} \quad (1.2)$$

Precision:

$$Pr = \frac{TP}{TP + FP} \quad (1.3)$$

F-measure:

$$Fm = 2 \frac{Pr}{Re + Pr} \quad (1.4)$$

False-positive rate (FPR):

$$FPR = \frac{FP}{FP + TN} \quad (1.5)$$

False-negative rate (FNR):

$$FNR = \frac{FN}{TP + FN} \quad (1.6)$$

Percentage of wrong classifications (PWC):

$$PWC = 100 \frac{(FN + FP)}{(TP + FN + FP + TN)} \quad (1.7)$$

Balanced-Accuracy (BAC):

$$BAC = \frac{Re + Sp}{2} \quad (1.8)$$

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. 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.

Mean True Positive Ratio (mTPR) This metric measures the ability of the object detection method to successfully include and classify the real moving object into the detected ROIs. It is calculated using :

$$mTPR = \frac{mTP}{mTP + mFN} \quad (1.9)$$

Where mTP represents the mean number of the pixels correctly classified as being in the moving region over the entire sequence. Similarly, mFN is the mean number of pixels wrongly classified as non-moving object pixels. Failing to classify a moving block as part of a moving region leads to failure to transmit the corresponding infor-

mation which will be missing at the destination.

Image Quality Metrics (PSNR, SSIM, MS-SSIM, VIF and BRISQUE) The frames quality evaluation is performed using the peak signal-to-noise ratio (PSNR), Structural similarity (SSIM), Visual information fidelity (VIF) and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). The PSNR is defined in dB as :

$$\text{PSNR} = 10 \log_{10} \frac{(2^n - 1)^2}{\text{MSE}} \quad (1.10)$$

Where n is the pixel depth and MSE is the mean square error computed, for $N \times M$ image, using :

$$\text{MSE} = \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M (x_{i,j} - y_{i,j})^2 \quad (1.11)$$

Where $x_{i,j}$ defines the original pixel value and $y_{i,j}$ the new pixel value after compression. The SSIM metric is defined by the following equation :

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + \theta_1) + (2\sigma_{xy} + \theta_2)}{(\mu_x^2 + \mu_y^2 + \theta_1)(\sigma_x^2 + \sigma_y^2 + \theta_2)} \quad (1.12)$$

Where μ_x and μ_y are the local means, σ_x and σ_y are the standard deviations and σ_{xy} is the cross-covariance for images x and y sequentially. θ_1 and θ_2 are two numerical stabilizing constants. We adopt for BRISQUE, VIF and MS-SSIM metrics the definitions presented in [111], [112] and [113] respectively. BRISQUE has the particularity of allowing a no-reference image quality assessment that has been widely used as a video quality assessment (VQA) metric for video surveillance systems. In [114], authors have shown the importance and the need to know about the minimum video quality required to ensure the efficient performance of AI algorithms. A high BRISQUE score indicates lower video quality.

ROI-based video coding strategy for low bitrate surveillance

2.1 Introduction

This chapter presents the design of an efficient ROI-based video coding strategy for wireless surveillance systems. The proposed method is a fusion of three key techniques: edge detection, frame differencing, and the sum of absolute differences, which is further optimized through the application of morphological operations. The frame blocks are then categorized into moving and stationary components through thresholding, thus enabling the compression and transmission of only the moving elements in an object-based video coding scheme. The results demonstrate the efficiency of this proposal in terms of precise detection and data gain.

2.2 Proposed method

2.2.1 Edge Detection

Edge detection has been widely used in moving object detection [115] [92] because of its simplicity and efficiency. There are many edge detection techniques in the literature

like Sobel [116], Canny [117] [118], Prewitt [119] and many others [120]. However, the edge detection technique suffers from false positives since the edge information does not contain only the edge of the moving object but also those of stationary objects. This problem could be solved by using a post-processing step [121].

2.2.2 Sum of Absolute Differences (SAD)

This technique has been introduced with the motion estimation techniques used in the new video standards [122]. It is based on FD where a pixel-wise difference between two frames is performed. Characterized by its simplicity, FD has been widely used in the literature for moving object detection [123] [98] [124]. The sum of the absolute difference of two consecutive frames, based on non-overlapping blocks of size $w \times w$ pixels, is given by :

$$SAD(x, y) = \frac{1}{w^2} \sum_{u=0}^{w-1} \sum_{v=0}^{w-1} D(wx + u, wy + v) \quad (2.1)$$

Where $x \in 0..M/w - 1$ and $y \in 0..N/w - 1$ are block indices and D is the difference between two consecutive frames of size $M \times N$ computed using :

$$D(i, j) = |F_n(i, j) - F_{n-1}(i, j)|, \quad i \in 0..M, j \in 0..N$$

This leads to an activity map of w^2 times less than the input frame size.

2.2.3 2-D Rank Order Filter

A rank order filter is a class of filters where the value of the center pixel is replaced by the appropriate order among its neighbors. The value of the center of each window is replaced by a chosen order value. A fast algorithm of this filter with $O(N)$ complexity was presented in [125]. In our experiment, we use the maximum (64^{th} for 8×8 kernel) order which is referred to as the maximum rank order filter. An example of the effect

of the filter on the activity map is presented in Figure (2.1) where the maximum rank order filter is applied to the scores of a sample map of size 8×8 .

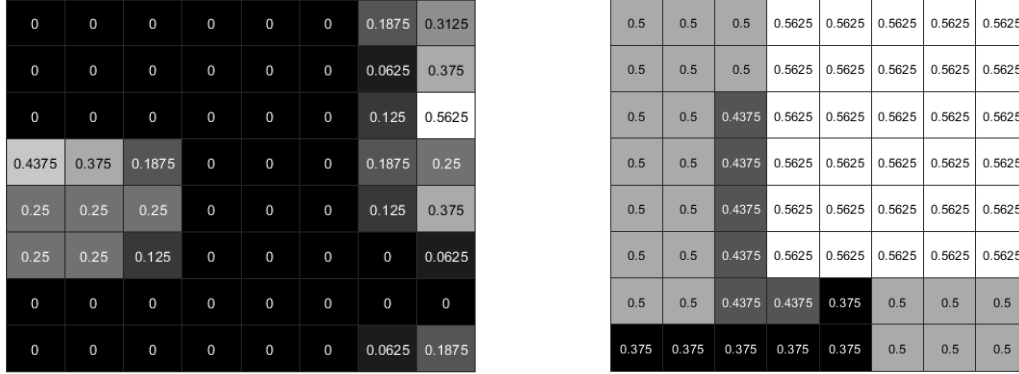


Figure 2.1: Impact of 2-D Rank Order Filter: from Right to left: Before, After

2.2.4 Fast Global Smoother

FGS is a type of Gaussian filter that was proposed by Dongbo et al. [126]. It is a global smoother that performs a spatially inhomogeneous edge-preserving image smoothing. FGS is recommended in our work since it is (i) able to eliminate noise while preserving edges and (ii) computationally effective with few arithmetic operations as it uses five multiplications and one division for one pixel.

The proposed ROI detection technique aims to apply edge detection of the incoming frame and the previous frame. We adopt the Sobel operator [127] as an edge detector because of its efficiency and low overhead. After the edge detection is performed, an absolute difference is made between the two edge maps.

Since the obtained edge map contains few details on the movement made between the two frames, enhancing the scores of the zone of moving regions is needed. To do so, we perform the SAD algorithm on the resulting map. This last step is important in two ways. On the one hand, summing up pixels in non-overlapping blocks allows getting a map in which regions with high movement will have significant scores while those

with no to low movement will get low scores values, which will make the extraction of the ROI easier and more accurate. On the other hand, it reduces the moving map to w^2 times its original size, which benefits the complexity part since the only reduced map will be considered for filtering, thresholding and storing in the sequent steps.

After that, a maximizing rank order filter is applied to the map to enhance the scores of the regions where there is significant movement and reduce the score where there is less movement. Next, the FGS filter is applied to the map to smooth it and create a homogeneity between the neighboring movement regions.

The last step of the proposed strategy is to extract the binary mask by performing a thresholding operation. The blocks labeled as moving blocks are concerned with compression and transmission to the destination while non-moving blocks are deleted. The ROI detection algorithm is presented in Figure (2.2). To eliminate the error propagation, we propose to transmit the whole frame each time a Group Of Pictures (GOP) is reached. The complete video coding strategy is presented in Figure (2.3).

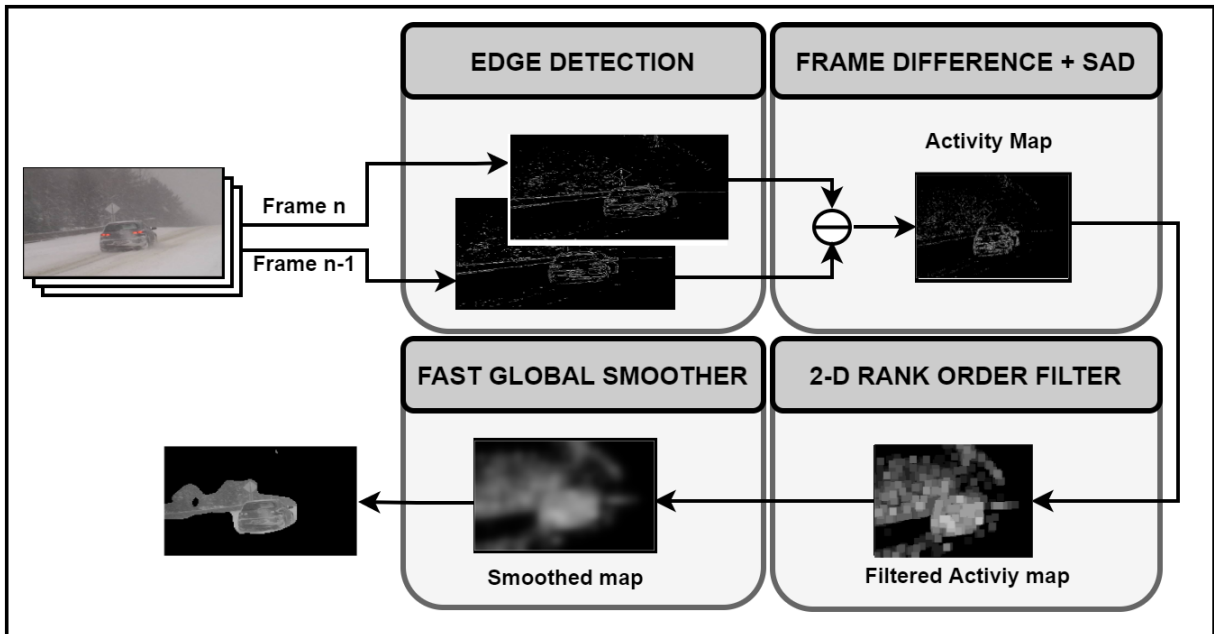


Figure 2.2: Block diagram of the proposed ROI Detection.

To validate the proposed method, selected videos from different datasets are used.

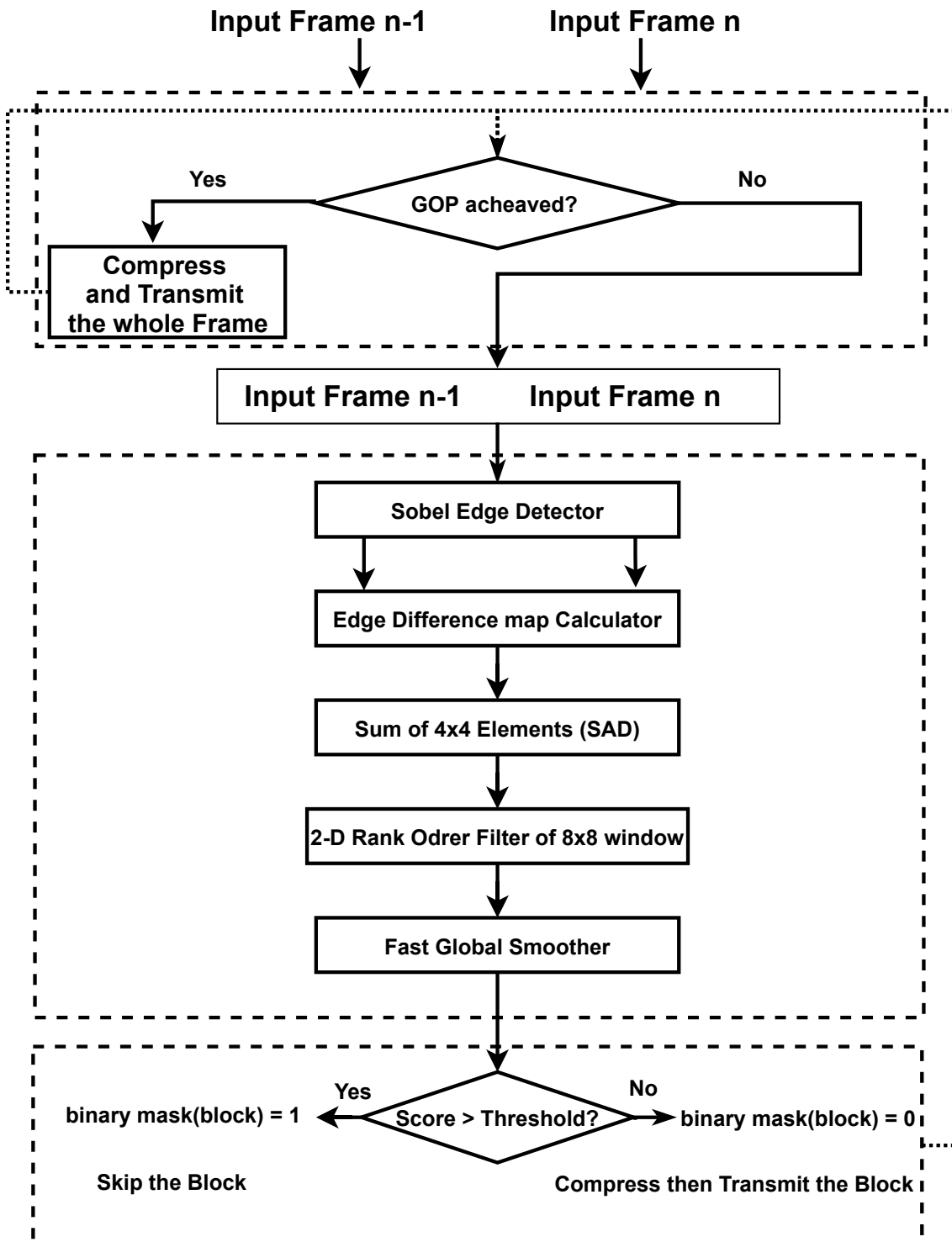


Figure 2.3: Block diagram of the proposed video coding strategy.

We have used the highway video from CDnet 2014 dataset (320×240) [128]. Freeway (316×236), peds (232×152), rain (308×228), traffic (378×282) videos from JPEGS dataset [129]. Traffic video (160×120) and Atrium video (640×360) from MATLAB and M-30 video from [130]. Further details are shown in Table 2.1.

Table 2.1: Details of the used dataset

Video Name	Dataset	Size	# frames
highway	CD net 2014 [131]	320×240	1700
peds	UCSD [132]	332×152	170
freeway	CAVIAR [133]	316×236	44
Rain	CAVIAR	308×228	229
Traffic	MATLAB	160×120	120
Traffic2	MATLAB	640×360	190
Traffic3	CAVIAR	378×282	190
Atrium	MATLAB	640×360	600
M-30	GRAM-RTM [134]	640×360	531

We analyze the detection efficiency in terms of visual object detection masks. The binary mask is constructed using ones and zeros by labeling ROI blocks with ones (On) and non-ROIs with zeros (Off). All the simulations are performed on MATLAB 2020a software running on a Quad-core i7 2.5Ghz laptop with GeForce GT 750M

Table 2.2: Used parameters and values for the simulation

Step	SAD		FGS		ROF	
Parameter	W	m	σ	λ	Percentile (p)	Wind. (K)
Value	4	5	0.035	30	100	8

2.3 Qualitative Results

We evaluate the qualitative results in terms of the visual binary mask of the selected blocks from the ROI detection. Table 2.3 represents the extracted regions where a high

and important movement has occurred.

It is depicted from the tests on the used dataset, in Table 2.3, the ability of the proposed algorithm to extract the ROI and include all the moving objects in the ROI. Including the extra edges of the moving region in the ROI will contribute to ensuring no missing visual information at the destination. And also, small and big objects are entirely detected. Also, we observe that the algorithm has a high sensitivity to detect near and far objects from the camera. Independently from the Field of View (FoV) of the used camera. The obtained results in terms of visual masks confirm that the algorithm is a good candidate for many pre-processing applications such as -surveillance systems where moving object-based video coding is needed or to reduce motion estimation cost [135] [136]. Also, it shows good performances for applications where moving multi-person and multi-object tracking is needed [137].

Table 2.3: Visual binary mask for the ROI detection.





















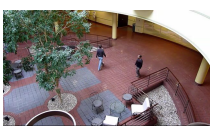









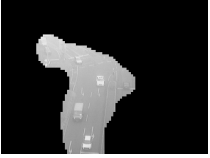

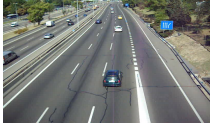



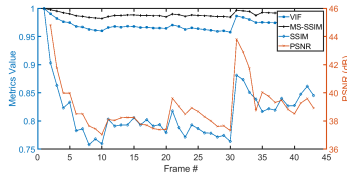
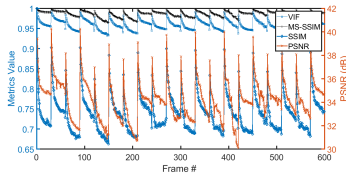
Sample	Original Frame	Mask	Moving Blocks	Reconstr. Frame
Highway #140				
Freeway #18				
Peds #44				
Rain #85				
Traffic #100				
Atrium #330				
Traffic2 #216				
Traffic3 #100				
M-30 #100				

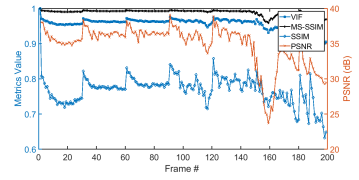
Table 2.4: PSNR, SSIM, MS-SSIM and VIF results for the used dataset



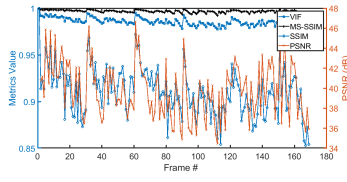
(a) freeway



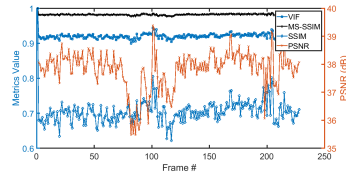
(b) atrium



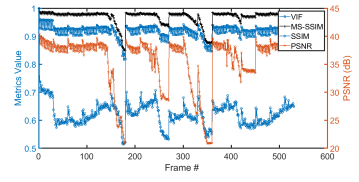
(c) highway



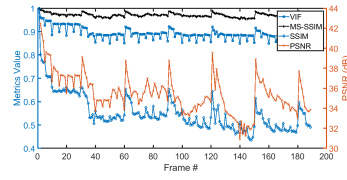
(d) peds



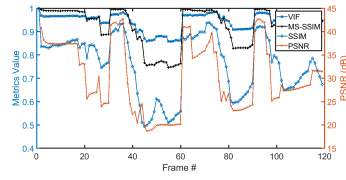
(e) rain



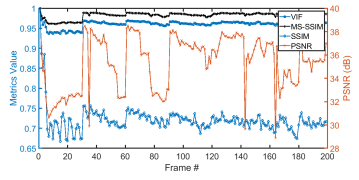
(f) Traffic2



(g) Traffic3



(h) Traffic



(i) M30

2.4 Quantitative results

We evaluate the quantitative results using multiple quality parameters. We calculate the PSNR, SSIM, MS-SSIM [113] and VIF [112] between the original frame and a reconstructed frame. A reconstruction of the frame is done by replacing the blocks in the previous original frame with the blocks detected in the new frame using block indexes. This method allows a complete evaluation of the quality of the reconstructed images and evaluates the performances of the detection.

Table 2.4 shows the quality metrics for different data sets used in this experiment. The SSIM and MS-SSIM values indicate the structural similarity between the original frame and the reconstructed frame. It is clearly shown that SSIM and MS-SSIM keep high values for almost all the sequences with values not less than 0.92 from MS-SSIM and 0.95 and higher for SSIM. The proposed strategy guarantees the reconstruction of the original frame each time a GOP is reached, by eliminating the temporal error propagation effect.

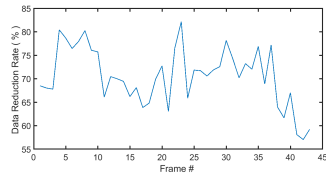
2.4.1 Data Reduction

To evaluate the performances of the proposed algorithm in terms of data reduction, we calculate the ratio of the data to be transmitted in our scenario to the total ratio where no pre-processing operation is performed (standard techniques). The gain in data is shown in Table 2.5.

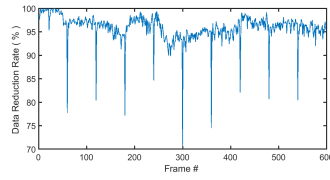
2.4.2 Data Saving over the used Dataset

Considering the scenario where non-active blocks are dropped. And only the blocks of the ROI are sent to the destination. We evaluate the data reduction in terms of the ratio of the blocks sent to the destination to the total ratio (the complete frame in classical scenarios like M-JPEG coding).

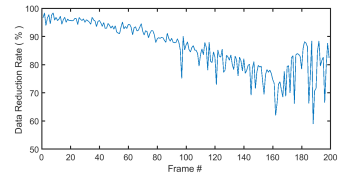
Table 2.5: Ratio of data reduction using the proposed strategy



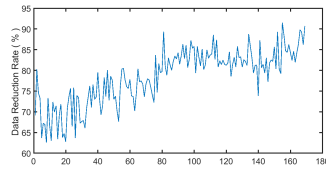
(a) freeway



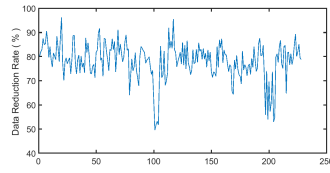
(b) atrium



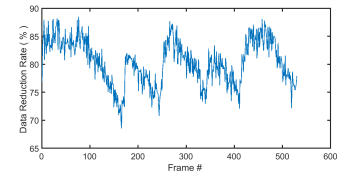
(c) highway



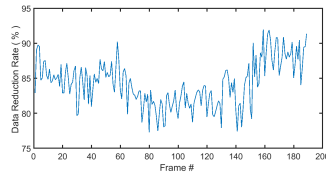
(d) peds



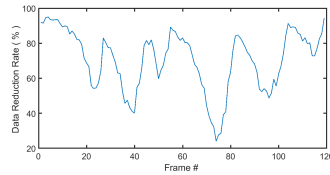
(e) rain



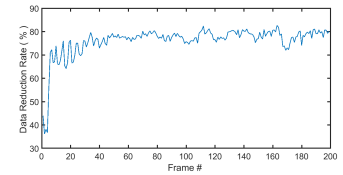
(f) Traffic2



(g) Traffic3



(h) Traffic



(i) M30

Figures in Table 2.5 show the ratio of data saving for the used test datasets. The sequences with small moving objects show high data saving as shown in figure (b) of the atrium video. We see that the data saving for the atrium video achieves nearly 98% of the saving since moving objects are human bodies with a high FoV of the used camera. While most of the other sequences achieve also high savings reaching 80% in mean for all videos.

Results show the high efficiency of the method to reduce data to be transmitted and thus, reduce the needed bitrate and needed transmission energy which is relatively very expensive in WMSN-based surveillance systems.

2.4.3 Comparison with Standard methods

In a classical scenario like the MJPEG compression technique, block-based compression for all the blocks is done for each frame. While, for the proposed strategy, only some blocks are compressed and transmitted. Table 2.6 shows the difference in terms of the number of blocks needed to be sent using our ROI-based technique and classical techniques. The table also shows the high amount of data reduction and radio energy dissipation saving.

Table 2.6: Mean number of blocks to be transmitted for each strategy

Sequence name	Sequence size	ROI-based (ours)	Classical approach	Saving(%) (wr. to classical)
Traffic2	640x360	4695	14400	67.4%
Atrium	640x360	589	14400	96%
Highway	320x240	1345	4800	72%
freeway	316x236	530	4661	88.6%
peds	232x152	719	2204	67.4%
rain	308x228	2132	4389	51.5%
traffic	378x282	1768	6662	73.5%
traffic3	160x120	428	1200	64.3%

2.5 Limits in terms of visual Quality

Error propagation over the successive frames is due to an error in the detection of the ROI. This problem affects the visual quality of the reconstructed frame. Figure (2.4) explains the effect of imprecise ROI detection. To solve this problem we adopt the following strategy with the proposed algorithm: after reaching a predefined GOP value, the algorithm sends all the frame blocks. This solution is efficient in breaking the continuous error propagation over successive frames. And it helps save an acceptable visual quality as shown in the results in Table 2.4.

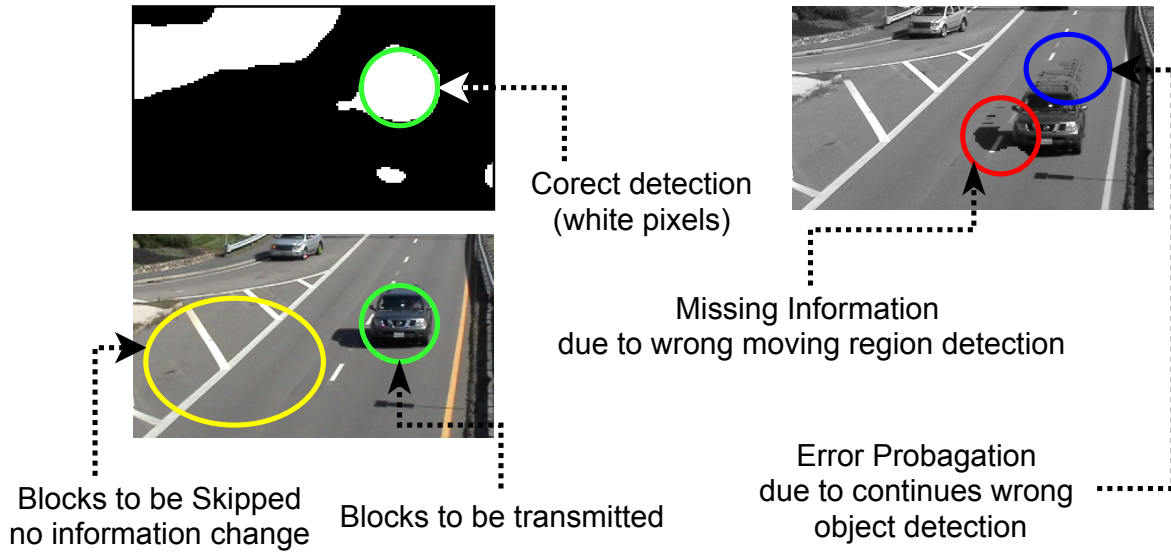


Figure 2.4: Example of the effect of wrong detection on the visual results degradation and error propagation. Frame #150 of traffic2 sequence)

2.6 Execution time

Table 2.7 reports the execution time in ms/frame for different frame sizes. The table shows the low time complexity of the method and its ability to achieve a high frame rate when implemented in low-cost surveillance platforms.

Table 2.7: Execution Time in milliseconds for different frame size

sequence size	execution time (ms) per frame
640x360	124
320x240	41
316x236	44.1
232x152	36.8
308x228	43.2
378x282	56
160x120	32.8

2.7 Conclusion

In this chapter, we presented an innovative object-based video coding strategy for low-bitrate surveillance systems. The proposed approach detects ROI blocks by utilizing edge features between consecutive frames and implements error correction by signaling complete frame updates at fixed GOP intervals. Our results demonstrate the efficacy of Sobel edge detection in identifying changes between frames and the effectiveness of SAD in refining edge features to accurately locate ROI. Our strategy delivered substantial bandwidth savings and transmission energy reduction (ranging from 51.5% to 96%) compared to traditional coding methods while preserving high-quality frame reconstruction. However, the method's reliance on Sobel edge detection, with its relatively high computational cost, and the absence of analysis regarding the accuracy of the detection method and its energy consumption under specific wireless sensor node conditions, were noted as weaknesses. These limitations will be addressed in the upcoming chapter through the proposal of the BIRD algorithm.

An Efficient Low Complexity ROI Detection for Video Coding in WVS

3.1 Introduction

In this chapter, we present a novel approach to tackle the challenge of balancing accuracy and energy efficiency in video coding for wireless visual surveillance (WVS) systems. Our proposed ROI detection algorithm serves as a pre-encoder and is designed to strike a balance between detection accuracy and computational complexity. To accomplish this, we create an activity map by measuring the motion activity of each block between consecutive frames. The map scores are then processed using a combination of a fast Gaussian smoother and a rank-order filter for improved accuracy. Our algorithm only encodes and transmits blocks that contain motion, resulting in significant energy and bitrate savings of nearly 90% and 98%, respectively. The efficacy of our approach has been thoroughly evaluated using key performance metrics, such as TPR attaining a sensitivity of 80.84%. The findings show that the BIRD algorithm outperforms other state-of-the-art methods in terms of accuracy while maintaining low computational overhead.

3.2 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 [63]. To overcome the low accuracy of pixel-based detection of the frame

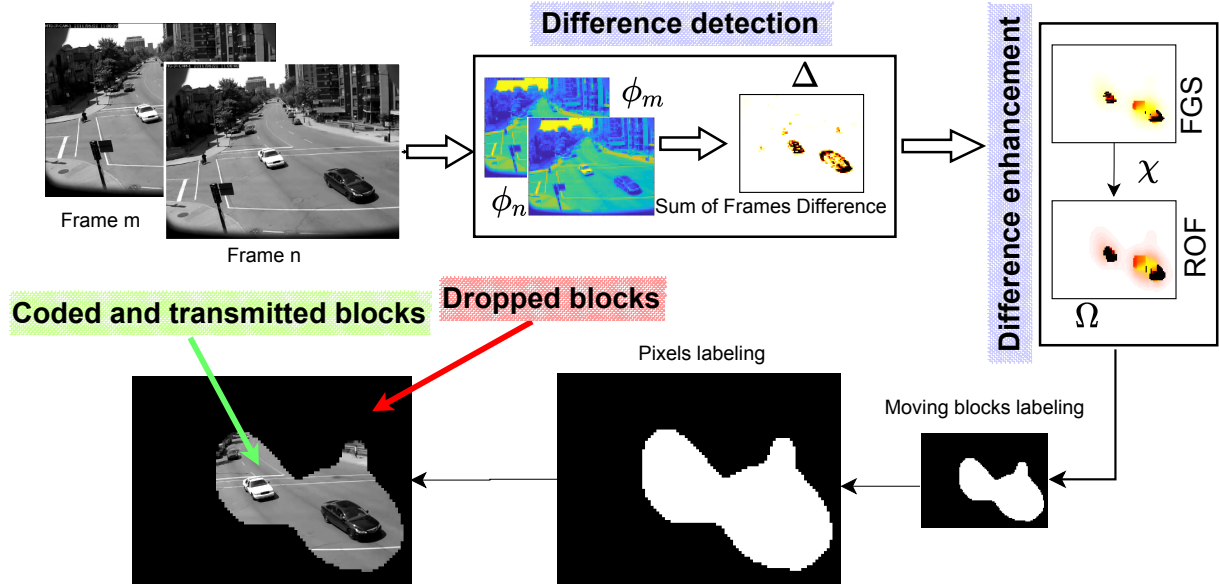


Figure 3.1: Block diagram of the proposed algorithm (BIRD)

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.

3.2.1 Difference Detection

Let ϕ_n and ϕ_m be the intensity map of the frames F_n and F_m of the size $M \times N$. Based on the SAD technique [138], the summation of the non-overlapping blocks of size 8×8

for F_n is provided by Equation 3.1

$$\phi_n(x, y) = \frac{1}{w^2} \sum_{u=0}^{w-1} \sum_{v=0}^{w-1} F_n(wx + u, wy + v) \quad (3.1)$$

While for the frame F_m , ϕ_m is calculated using Equation 3.2

$$\phi_m(x, y) = \frac{1}{w^2} \sum_{u=0}^{w-1} \sum_{v=0}^{w-1} F_m(wx + u, wy + v) \quad (3.2)$$

Where $x \in 0 \cdots M/w - 1$ and $y \in 0 \cdots N/w - 1$ are block indices. The resulting intensity maps ϕ_n and ϕ_m are w^2 times less than the input frame size F_n . To create the activity map Δ , the SAD operation is completed by computing the absolute difference between the two intensity maps, as in Equation 3.3

$$\Delta(w, y) = |\phi_n(x, y) - \phi_m(x, y)| \quad (3.3)$$

In view of this, the scores in Δ indicate the level of activity created between the two frames. The blocks that contain high movement are represented by high score values in Δ , 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 (3.1).

3.2.2 Difference Enhancement

To avoid the false negative problem and improve the accuracy, an enhancement of the scores of Δ is needed. We propose the combination of a smoothing and rank maximization of Δ . Therefore, we propose to take the advantage of both the efficiency and rapidity of the Gaussian smoother FGS [126].

As depicted in Figure (3.1), FGS is applied on the Δ map to smooth the details and noisy part resulting from the SAD operation. Contrary to the convolution filters, FGS

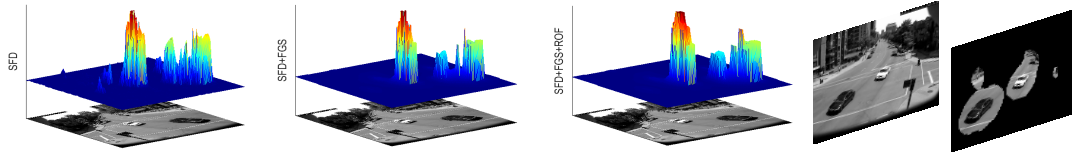


Figure 3.2: FGS eliminates unnecessary activities and ROF enhances the non-zeros scores prior to thresholding

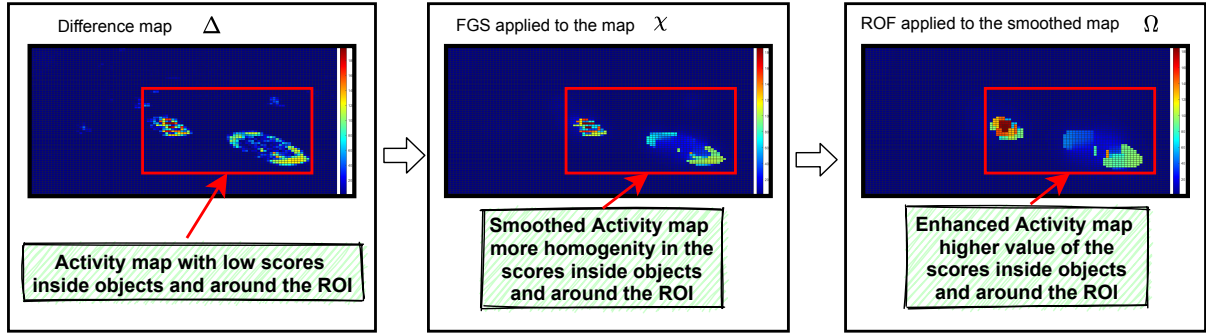


Figure 3.3: Impact of the combination of FGS and ROF on the ROI classification

is characterized by a low complexity and rapidity estimated to be over 30 times faster than other filters. FGS uses a parameter σ to control the variance around the mean value and another parameter λ to define the amount of regularization during filtering.

Subsequently, the resulting smoothed map (χ) is filtered by the maximum rank order filter (ROF). The ROF belongs to a class of filters easy to implement [139]. 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 [67]. 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 Ω . Figure (3.3) illustrates the impact of the used filters to enhance the ROI classification performances while Figure (3.2) summarizes the impact of each filter as used in this order.

The binary mask is then created by comparing the Ω 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 3.1 further summarizes the above steps.

Algorithm 3.1: The Proposed BIRD algorithm

Input:

m selected previous frame
 N SAD blocks size
 K ROF window size
 p rank order of the ROF
 T threshold value
 λ regularization of FGS
 σ 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 (3.1)(3.2) and (3.3);
 $\Delta \leftarrow SAD(F_n, F_m)$;
 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 Ω **do**
 if $Score(x, y) \geq T$ **then**
 Set $mask(block) \leftarrow 1$;
 Set $block_{ind} \in S_a$;
 else
 Set $mask(block) \leftarrow 0$;
 Report ROI mask to encoder ;
 Report $block_{ind}$ vector to receiver;

3.3 Results and Discussion

To validate the proposed method, we present the Change Detection 2014 Dataset (CDnet) [128] 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.

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

3.3.1 Parameters and experimental conditions

The experimental values for each used parameter are summarized in Table 3.1.

Table 3.1: Used parameters for the conducted simulations

Step	SAD		FGS		ROF
Parameter	N	σ	λ	p	K
Value	8	0.05	30	100	4

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

We use the metrics defined in Section (1.6) in chapter 1. Among those 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.

3.3.2 Performances of BIRD over the CDnet 2014

Table 3.2 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 3.2 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 [67].

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 3.3 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.

3.3.3 Comparison with other techniques

Table 3.4 shows the overall results of our method on CDnet 2014 dataset compared with the state-of-the-art techniques namely, KNN in [140], GMM in [141], KDE in [142], Mahalanobis Distance and Euclidean Distance techniques presented in [45] and another GMM-based technique in [143]. 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.

Table 3.2: Samples of ROI extraction mask results








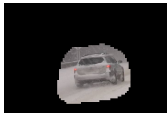















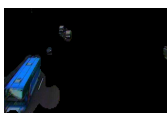
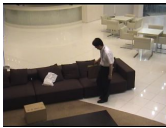


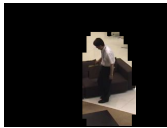
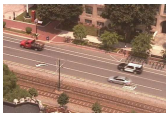







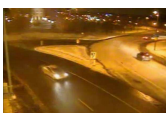



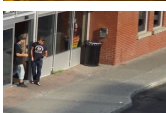



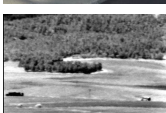



Sequence	Original	ground-truth	mask	ROI
Highway #1475				
SnowFall #2784				
Pedestrians #476				
Blizzard #1406				
WinterDriveway #1860				
tunnelExit #2329				
Sofa #1185				
PTZ #1240				
Park #250				
NightVideo #1300				
Busstation #400				
Turbulence0 #2045				

Table 3.3: Detection results of the proposed algorithm over CDnet 2014 dataset

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

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 3.3, 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.

3.3.4 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 [144]. The second method uses block-based object detection [145] same as our proposed method.

Table 3.4: Comparison of BIRD with classical techniques over CDnet 2014 dataset

Technique	Recall	Specifi.	FPR	FNR	PWC	F-Meas.	Precision
KNN [140]	0.6650	0.9802	0.0198	0.3350	3.3200	0.5937	0.6788
GMM1 [141]	0.6846	0.9750	0.0250	0.3154	3.7667	0.5707	0.6025
KDE [142]	0.7375	0.9519	0.0481	0.2625	5.6262	0.5688	0.5811
MahaD [45]	0.1644	0.9931	0.0069	0.8356	3.4750	0.2267	0.7403
GMM2 [143]	0.6604	0.9725	0.0275	0.3396	3.9953	0.5566	0.5973
EucD [45]	0.6803	0.9449	0.0551	0.3197	6.5423	0.5161	0.5480
BIRD	0.8084	0.8357	0.1642	0.1915	16.7115	0.1893	0.2678

As presented in Table 3.5, the BAC and recall metrics of BIRD show higher values than in [145] for most of the sequences. While [144] shows superior BAC and specificity values compared with BIRD and [145]. Results of BIRD are still very competitive to that of [144]. With an overall BAC of 82%, BIRD can ensure high detection accuracy of the moving object regions for different categories and conditions.

3.3.5 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 (3.4) plots the TPR against the FPR when varying the threshold value ($0 \dots 10$). 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, which means that more data is to be

Table 3.5: Category-wise comparison of BIRD to state-of-the-art on CDnet 2014 dataset

Category	Recall			Specificity			Blanced Accuracy		
	BIRD	Savas [145]	Cwizar [144]	BIRD	Savas [145]	Cwizar [144]	BIRD	Savas [145]	Cwizar [144]
Dynamic.	0.7593	0.6436	0.8144	0.9512	0.9962	0.9985	0.8553	0.8199	0.9064
PTZ	0.9662	0.7685	0.3833	0.6443	0.9977	0.9968	0.8053	0.8831	0.6901
BadWeat.	0.9208	0.5647	0.6697	0.8948	0.9985	0.9993	0.9078	0.7816	0.8345
Baseline	0.7619	0.6214	0.8972	0.9437	0.8213	0.9980	0.8528	0.7213	0.9476
CameraJ.	0.8504	0.4567	0.7436	0.6446	0.9788	0.9931	0.7475	0.7177	0.8683
Intermi.	0.4186	0.5547	0.8324	0.8603	0.9979	0.9911	0.6394	0.7763	0.9118
LowFram.	0.8161	0.5490	0.6659	0.7905	0.7464	0.9949	0.8033	0.6477	0.8304
nightVi.	0.9455	0.4593	0.4511	0.8374	0.9583	0.9874	0.8915	0.7088	0.7193
Shadow	0.8775	0.8365	0.8786	0.8500	0.9828	0.9910	0.8638	0.9097	0.9348
Thermal	0.7548	0.4650	0.7268	0.8894	0.9647	0.9949	0.8221	0.7148	0.8609
Turbule.	0.8216	0.7421	0.7122	0.8870	0.9883	0.9997	0.8543	0.8652	0.8559
Overall	0.8084	0.6056	0.6608	0.8357	0.9483	0.9948	0.8220	0.7770	0.8509

***bold:** the best category-wise, **red:** the best overall, **blue:** the second best

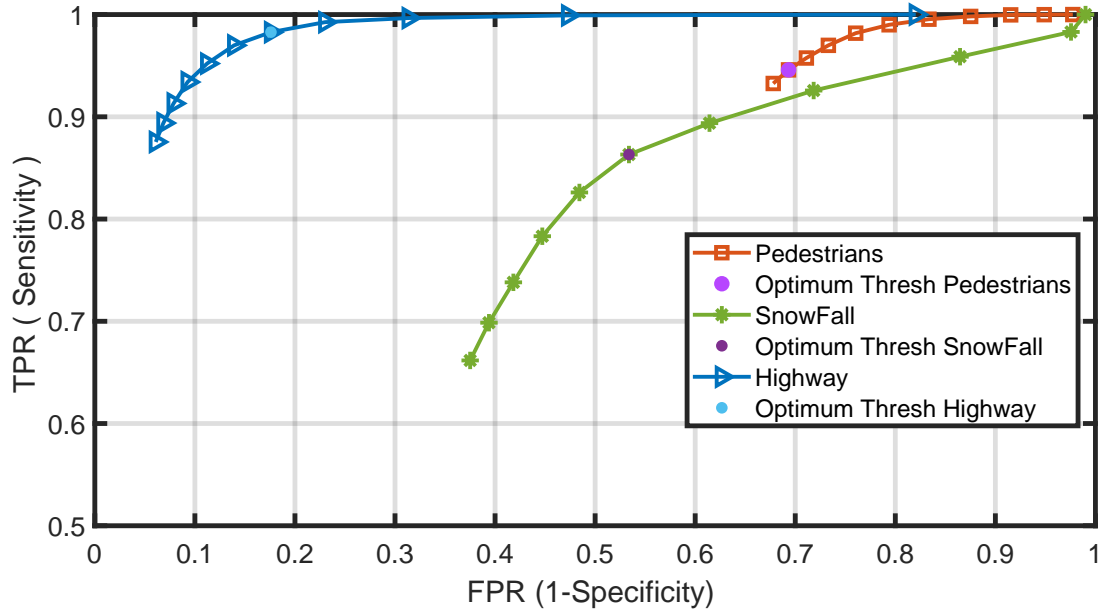


Figure 3.4: ROC curve and the optimum threshold for *pedestrians*, *Highway* and *Snowfall* sequences

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: $\min(\sqrt{(1 - \text{sensitivity})^2 + (\text{specificity} - 1)^2})$.

Figure (3.5) 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 destination. But, at the cost of higher energy and bitrate.

Table 3.6 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

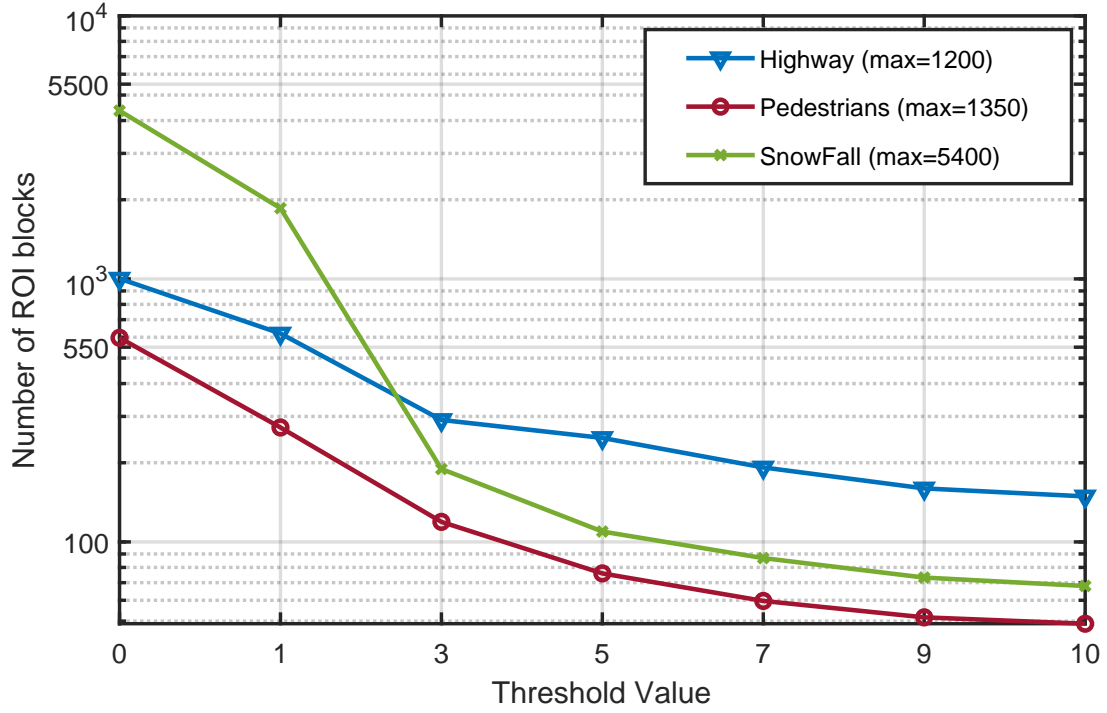


Figure 3.5: Number of blocks belonging to the ROI according to the threshold value

Table 3.6: Statistics of the energy gain under threshold variation

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

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 (3.4), for the *pedestrians* sequence, the optimum threshold for good detection accuracy is 9. Consequently, this threshold value saves about 96% of the processing and transmission energy compared to the CTA approach (see Table 3.6). An optimum threshold enables the optimum ratio of the activity blocks and could be used as a rate controller, which is a fascinating subject for future work.

3.3.6 Method Complexity

To evaluate the consumed energy on embedded sensor conditions, we consider what follows a sensor node equipped with an ARM Cortex M3 micro-controller [146]. Table 3.7 shows the processor characteristics. Using MATLAB 2020a and C++ running on a

Table 3.7: ARM Cortex M3 characteristics

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].

PC intel Core i7-2670QM 2.2Ghz, with 8GB RAM on Windows 7 OS, 2.6 ms to process one frame of 320×240 is recorded allowing processing of 384 frames per second (fps).

Table 3.8: Computational budget of each step of BIRD algorithm

Step	Operations	# of Operations	Energy consumption (mJ/Frame)	
			min ($Cyc_{div} = 1$)	max($Cyc_{div} = 12$)
SAD	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
$E_{detection}$	-	-	0.3723	0.6891

3.3.7 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 follow:

$$E_{Detection} = E_{SAD} + E_{FGS} + E_{ROF} + E_{Threshold} \quad (3.4)$$

The total computational budget of the method is presented in Equation 3.8. The number of operations for FGS is reported in [126] while the ROF budget is estimated using the mathematical model presented in Equation 3.5.

$$R = \frac{(K(K-1))}{2} \quad (3.5)$$

Where K represents the size of the sliding vector (K is set to 4 for the proposed method). 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 equal to

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 (8×8 , $16 \times 16 \dots$), a generalized model of the number of arithmetic operations should be presented. We present in Table 3.8 the number of operations for each step in terms of frame size (N, M) and block size (w). Table 3.8 also shows the energy budget of each step and the total energy budget of the BIRD. Table 3.9 shows

Table 3.9: Per-frame $E_{detection}$ cost of the method compared to state-of-the-art for size (240x320)

Method	Energy Budget (mJ/Frame)	
	min ($Cycles_{div} = 1$)	max ($Cycles_{div} = 12$)
MoG [141]		649.95
CS-MoG [147]		116.44
CoSCS-MoG [148]		125.96
EBSCAM [149]		3.4
FD		0.5069
BIRD (proposed)	0.3723	0.6891

a comparison of the energy budget of the proposed object detection method against state-of-the-art techniques for 240×320 , namely MoG [141], CS-MoC [147], CoSCS-MoG [148], EBSCAM [149] 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.

3.3.8 Energy dissipation for complete compression chain

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

$$E_{total} = E_{Detection} + E_{compress} \quad (3.6)$$

Where $E_{Detection}$ is the energy cost of the object detection part as presented by Equation 3.4, $E_{compress}$ is the energy cost of the compression part. For the calculation of $E_{compress}$, the model has been studied and provided in [150] under the same conditions.

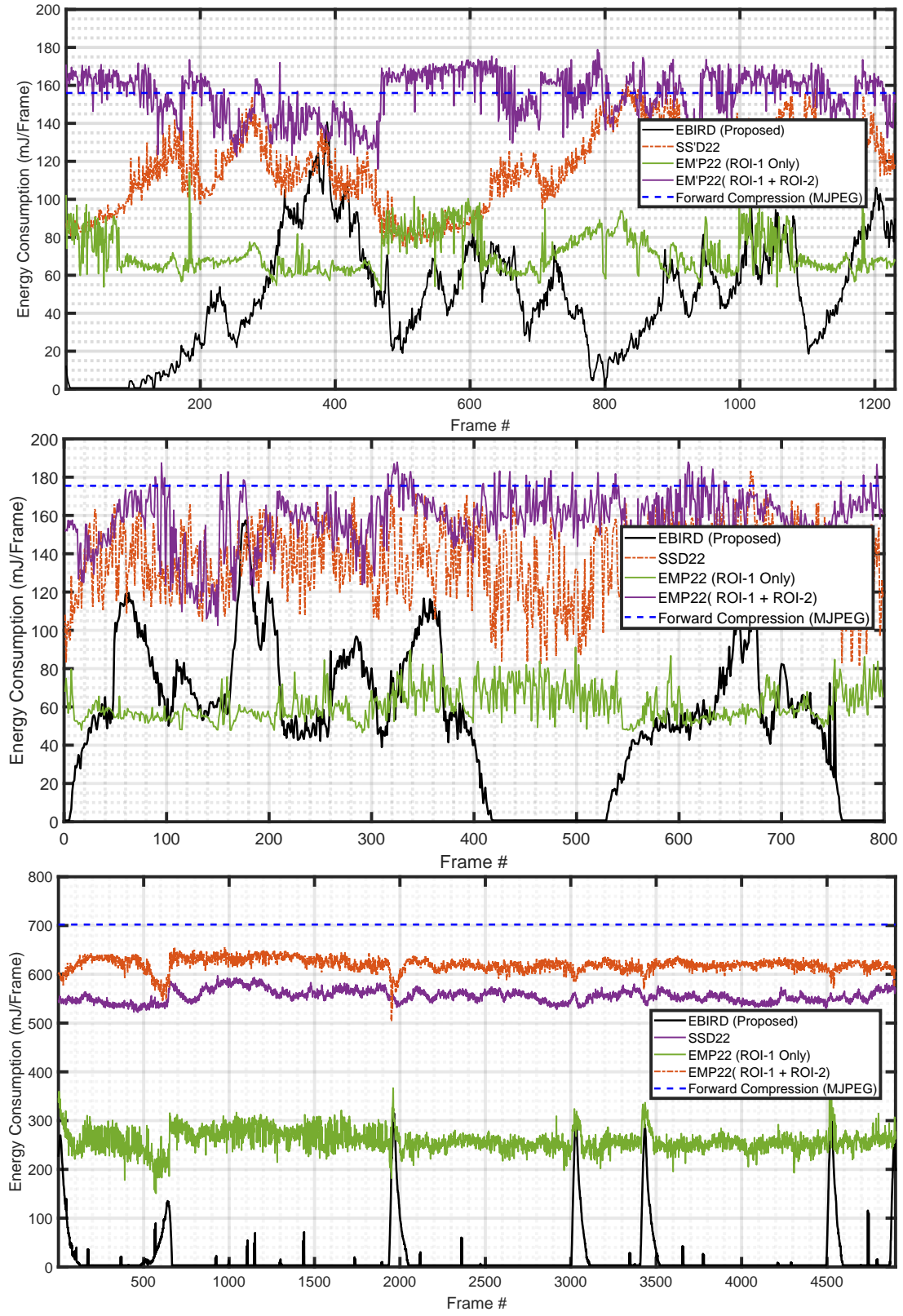


Figure 3.6: Per-frame energy dissipation of BIRD for *Highway*, *pedestrians* and *Snowfall*

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 [150] 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 8×8 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 $0.6891mJ/frame$.

Figure (3.6) illustrates the per-frame energy consumption of the proposed method compared to ROI-based compression methods, namely, [66] referred to as EMP'22, [65] 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.

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 [66] [65] 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 (3.6) 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.

3.3.9 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.

3.4 Conclusion

In this chapter, we proposed an energy-efficient moving region detection approach as a pre-encoder for WVS. The suggested approach is built upon a low-complexity SAD 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. This study focuses on the detection of the ROI as a binary classification (ROI/non-ROI). However, exploring the possibility of multi-class classification of the frame into multiple categories would be a valuable avenue of investigation. Such an approach has the potential to enable more precise content-aware coding and ensure higher QoS for specific regions within the frame. This will be the topic of investigation in the next chapter.

Multi-Threshold-based frame segmentation for content-aware video coding in WMSN

4.1 Introduction

In this chapter, we aim to push the boundaries of the current state-of-the-art by introducing a content-sensitive technique for video coding in wireless video surveillance with lower bitrates. Our proposed method utilizes ROI detection and coding within an adaptive compression paradigm. We detect high-activity regions in the video stream through automatic thresholding and then adjust MJPEG compression parameters based on the relative importance of each zone. We rigorously evaluate the effectiveness of the proposed method and find that it provides substantial benefits in terms of low bitrate and content awareness for wireless surveillance.

4.2 Proposed method

The proposed system is based on exploring the difference in edges of successive frames to detect the different moving regions. Edge Detection is applied using the Canny Operator [117]. Experiments show that the Canny operator includes weak and strong edges unlike other methods [151]. Thus, the results of the absolute difference will contain more edges, increasing the sensitivity of movement detection. For each frame, the result is a binary image that labels the edge pixels with ones and non-edge pixels with zeros. The sum of the absolute difference between the edges maps of the current and the previous frame gives high values to high-moving regions while values are lower for segments with few movements.

To enhance the obtained activity map, we perform morphological filtering by applying the 1D-Rank Order Filter (ROF) to the activity map. The ROF replaces each selected pixel with the max, the min, or the median value. The new value of the pixel is selected from sorting the neighbors of the pixel. In this work, we use the max ROF to remove impulse noise in the activity map and perform dilatation since it uses homogeneous maximization.

Next, the resulting activity map is smoothed using the Fast Global Smoothing filter (FGS). FGS is a fast Gaussian filter proposed in [126] that performs smoothing of the image. The FGS is used to distribute the enhanced values of the map (after application of the ROF filter) over the region, and remove holes inside region masks. The choice of sigma and lambda parameters is crucial in this work and directly affects the segmentation performances.

In the thresholding step, the binary map is divided into three segments based on the activity level: ROI-1 is the most important region and contains the moving object, ROI-2 is the second region and contains the region around the object, and ROI-3 is the last segment that contains the regions with the lowest priority and importance. The threshold is selected using the Otsu thresholding method [152]. By calculating

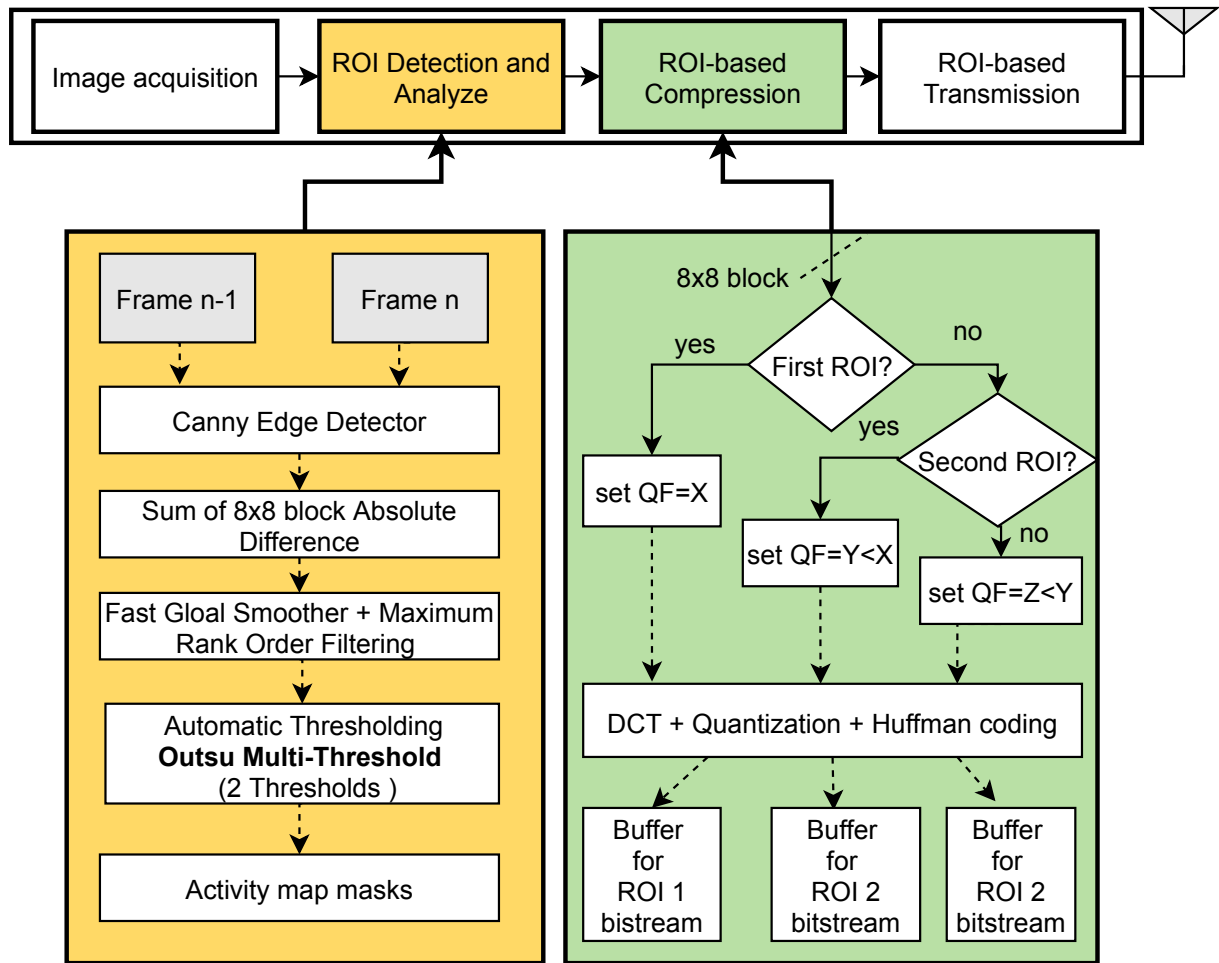


Figure 4.1: Proposed ROI-based video compression scheme

the histogram of the activity map, the Otsu thresholding method checks the existence of the proper thresholds (2 values in our case) to classify multi-regions based on the histogram distribution. We use MATLAB's built-in function *multithresh*, which implements Otsu's method of multilevel thresholding [152]. Figure (4.1) shows the details of the proposed method.

On the compression side, the compression is performed using the JPEG chain where 8x8 DCT is applied to each block. To give priority and the best delivery conditions to the activity region, the quantization factor (QF) is chosen based on the importance of the block. The blocks of the ROI-1 will be coded with high QF. ROI-2 and ROI-3 are coded using low QF to save bitrate and transmission energy.

4.3 Results and Discussion

Simulation is performed using MATLAB 2020a software running on a Quad-core i7 2.5Ghz laptop. Details are shown in Table 4.1.

Table 4.1: Parameters and methods used for each Step

Step	Edge Detector	SAD	FGS ROF Thresh.			JPEG		Classes QF			
Param.	Operator	Wind. size	σ	λ	n p	algo.	Comp. tech.	Entr. Cod.	X	Y	Z
Value	Canny	8	0.0530	4	100	Otsu	8-DCT	Huffman	90	50	10

4.3.1 Image segmentation results

Segmentation results are shown as a sample frame in Figure (4.2) where results of three regions are shown. We see the ability of the algorithm to accurately classify regions based on their movement and their importance.

Figure (4.3) Shows the impact of the ROI detection on the compression visual quality. The proposed method considers the moving object as the most important part that has to get the highest priority and the highest quality, the second segment with



Figure 4.2: Example shows the results of segmentation for the 3 regions, 1-Frame 85 form Hall sequence. 2-Mask of the 3rd ROI. 3-Mask of the 2nd ROI. 4-Mask of the 1st ROI (Th1=0.0224, Th2=0.0301)

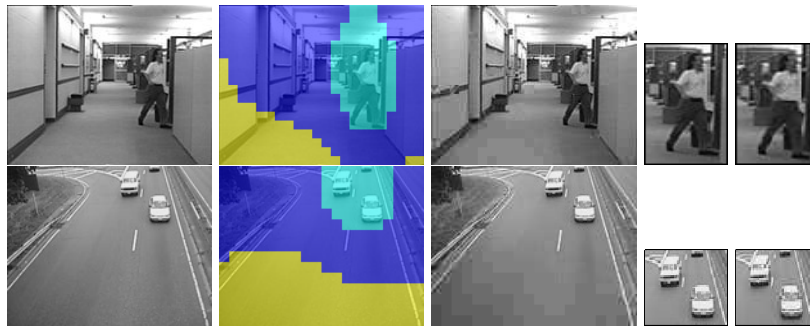


Figure 4.3: Results of multi-QF based coding. from left to right: 1- Original Frame 2- Segmentation Results 3- Decompression results (JPEG chain with ROI1: QF=90, ROI2:QF=50, ROI3:QF=10 - PSNR=33.9308, SSIM=0.7618) 4- ROI visual quality for same bitrate(proposed left, MJPEG right).

medium importance as a second priority and the regions with the least priority which are the non-moving regions.

4.3.2 Compression Quality Results

We Evaluate the effect of multi-class QF decision for the compression step on the Hall sequence (176x144) and Traffic sequence (160x120) using the PSNR and SSIM metrics. Figure (4.4) and Figure (4.5) show the PSNR results for 300 frames of Hall sequence and 120 frames of Traffic sequence for the proposed Multi-QF coding method using three levels QF(90,50,10) in comparison with MJPEG at QF=90. The MJPEG is chosen for comparison due to its low complexity compared to recent encoders and since it shows a large implementation in WMSN. It is shown that the PSNR value is lower for the case of multi-QF in comparison with MJPEG. The reduction is generally about 9dB for Hall and Traffic sequences. The PSNR is generally about 35dB to 31dB which is still

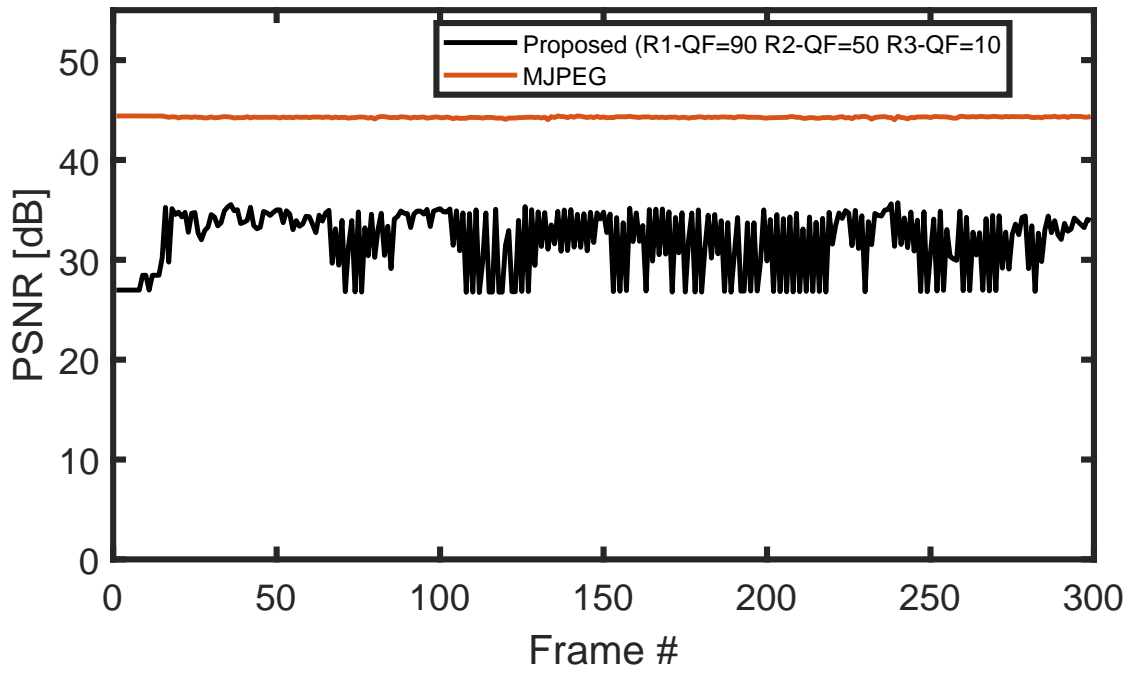


Figure 4.4: PSNR value of ROI-based coding compared to MJPEG for Hall seq.

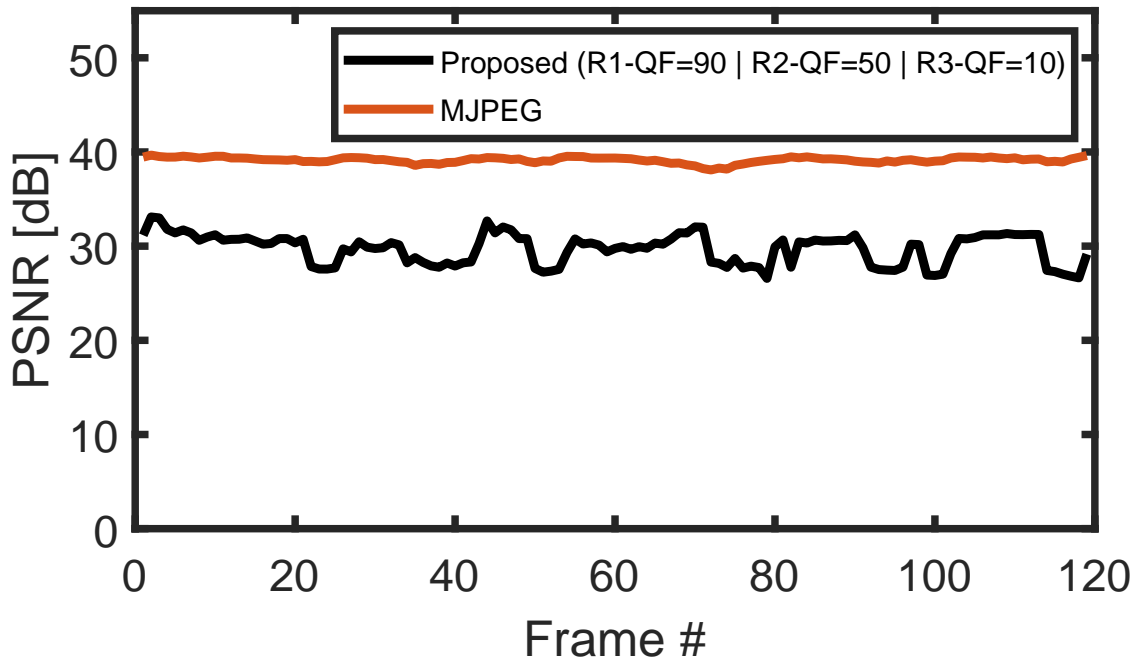


Figure 4.5: PSNR value of ROI-based coding compared to MJPEG for traffic seq.

very acceptable.

For the same bitrate, Figure (4.6) and Figure (4.7) show the quality of the delivery of the ROI-1 that contains the most important information of the frame. The delivery with about 37.32db and 39.40dB for Hall and Traffic sequences successively is better than MJPEG (31.70dB and 34.94dB) with a slight degradation of the whole video quality using Multi-QF for non-important regions (ROI-2 and ROI-3).

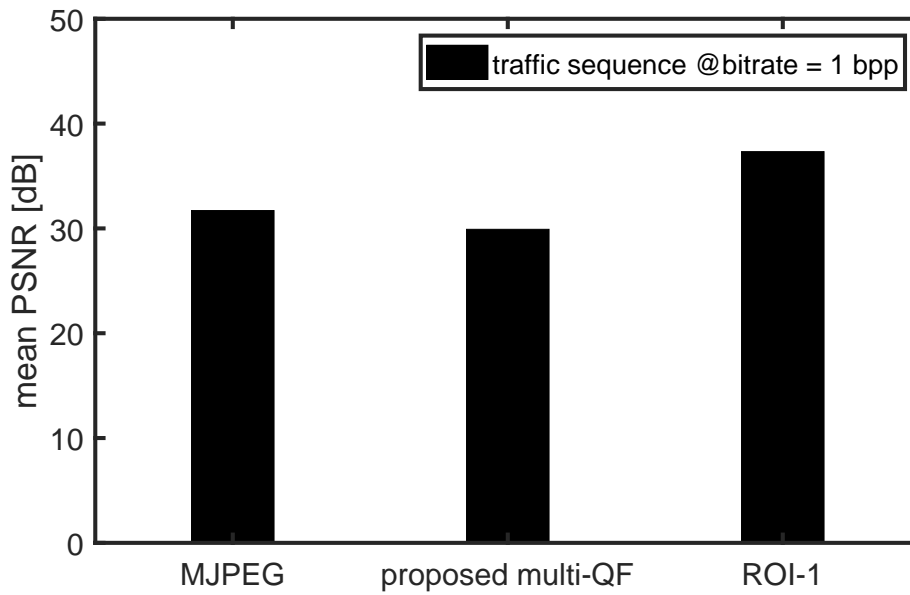


Figure 4.6: Mean PSNR of the whole frame and the ROI-1 for the proposed method compared to MJPEG for the same bitrate (hall seq.)

Figure (4.8) shows the SSIM results for the hall video. The reduction in terms of structural similarity is generally from 0.25 to 0.1 compared to MJPEG at QF of 90 (0.98).

4.3.3 Bitrate Results and Gain

Figures (4.10) and (4.11) highlight the large reduction of transmission bitrate (generally more than 50%) when the proposed method is employed against MJPEG. This leads to a reduction in bandwidth usage and, thus, less contention in the channel, a common problem in WSN.

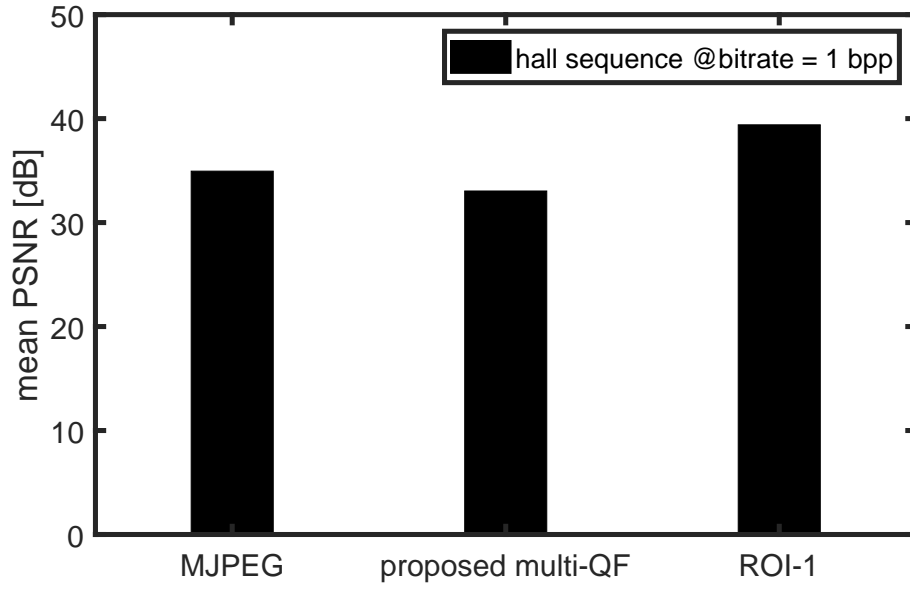


Figure 4.7: Mean PSNR of the whole frame and the ROI-1 for the proposed method compared to MJPEG for the same bitrate (traffic seq.)

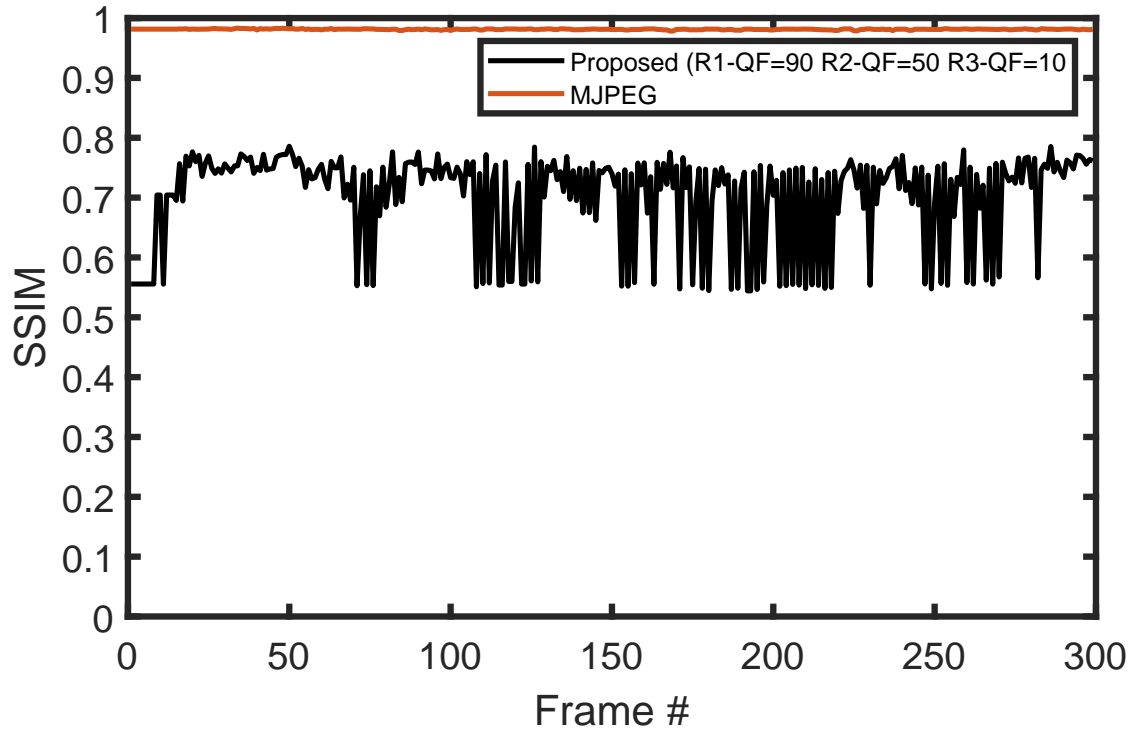


Figure 4.8: SSIM results of ROI-based coding compared to MJPEG for hall seq.

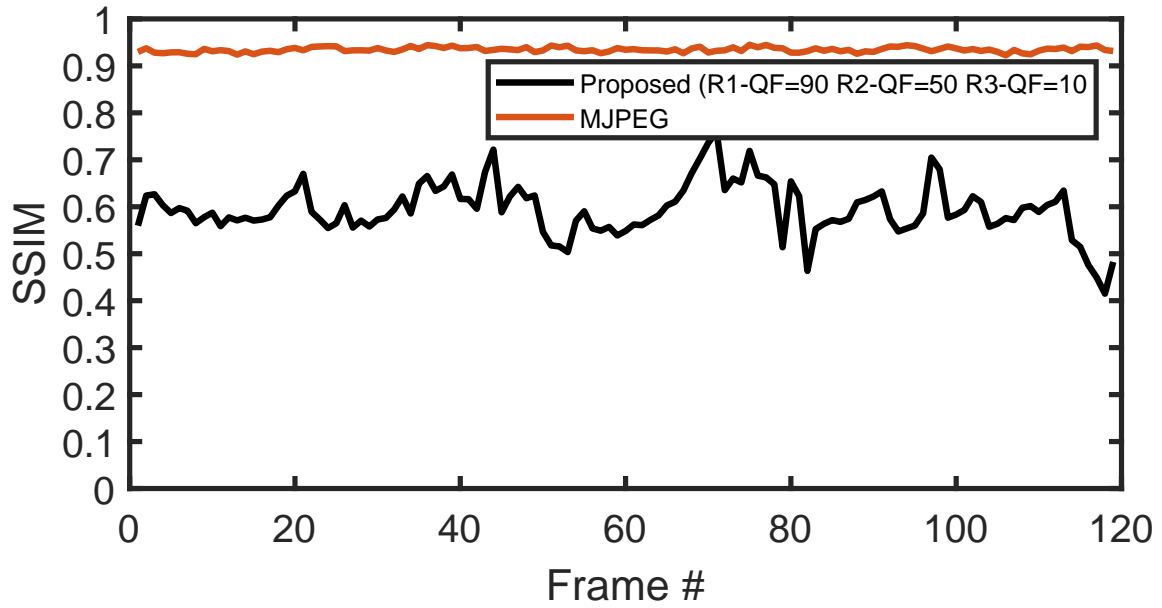


Figure 4.9: SSIM results of ROI-based coding compared to MJPEG for traffic seq.

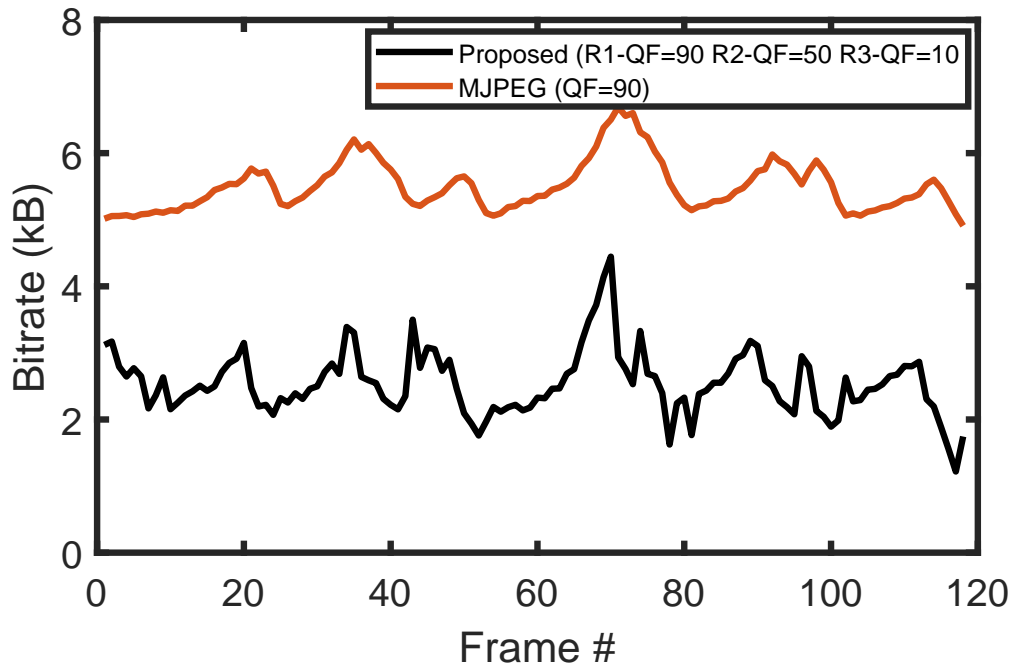


Figure 4.10: Required bitrate for proposed strategy against standard MJPEG for traffic seq.

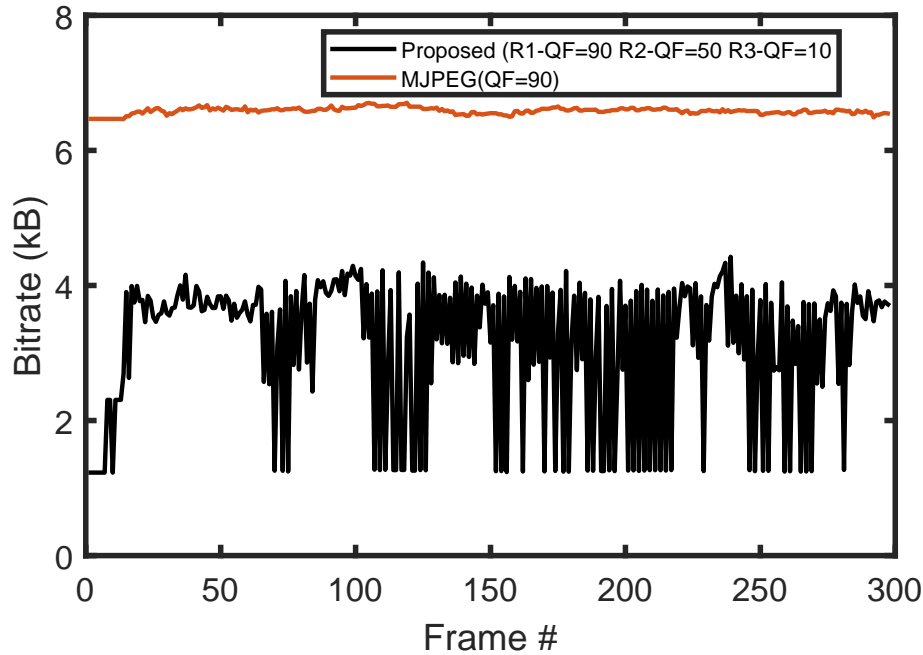


Figure 4.11: Required bitrate for proposed strategy against standard MJPEG for hall seq.

For scenarios such as multi-hop environments, in which other energy-constrained nodes would be relaying the frames, the energy saving of using the proposed method would increase with the number of hops since the bit-reduction will propagate across the network.

Using automatic multi-zone segmentation with automatic thresholding, the suggested method has led to good results in preserving the high quality of essential parts of the frames while reducing the bitrate. Another advantage of the proposed technique is its minimal complexity, which is tailored to the ATC paradigm for WMSN, as opposed to contemporary encoders, which are too complex for embedded sensor nodes.

4.4 Conclusion

In this chapter, a low-cost multi-threshold frame segmentation method is proposed and applied to video compression using a multi-quality factor based on moving re-

gion importance. The results of the proposed method show high visual quality for moving objects. A delivery guarantee of the essential blocks allows advanced tasks at the reception of the visual data. The proposed method also shows a high bitrate saving of more than 50%. The current limitations of the Canny Edge Detector and Otsu Thresholding method are their high energy costs. To overcome this challenge, a lower-cost detection technique employing simple arithmetic operations such as summation, which incurs minimal processing overhead, must be explored. In the following chapter, a novel lower-cost multi-class ROI detection technique is proposed. A comprehensive evaluation of the technique will be conducted on a larger dataset, with a deep analysis of computational complexity and energy consumption to address this weakness.

ROI-based video coding strategy for rate/energy-constrained smart surveillance systems using WMSNs

5.1 Introduction

Reducing the power consumption of the in-node processing and the required bandwidth while maintaining a high QoS is a challenging task. The difficulty increases when a smart task must be performed on the received video in the destination. In this context, this chapter proposes an energy-efficient video coding strategy based on a new and fast ROI detection method. The lightweight ROI detection method segments the frame into four regions. And the coding strategy aims to extract two different classes of the ROI for coding and transmission using variable quality levels based on their relevance. Furthermore, the strategy aims to exclude the regions of lower importance and any non-ROI that has insignificant movement. We assess the strategy's ability to perform object recognition tasks at the destination under quality degradation. The performance results using different datasets demonstrate a better trade-off between awareness of ROI quality, energy consumption and bandwidth savings for

the proposed strategy compared to other methods. This results in a 96% reduction of bandwidth and 93% reduction in energy for some sequences at the expense of a 1-4 dB decrease in PSNR when compared to the MJPEG standard. While the recognition accuracy of the YOLOv3 model at the destination outperforms the other techniques by about 4% to 22%.

5.2 Proposed S-SAD method

In an attempt to enable multimedia applications using constrained wireless networks, we propose an efficient and low-cost multi-level ROI detection scheme prior to the compression and transmission phases. Depending on the importance of each level, a decision is made on whether or not to transmit the corresponding region. If so, a compression with an appropriate quality coefficient reflecting the importance of the region is applied on its blocks prior to transmission. Transmitting only a subset but relevant information of each captured image saves energy and bandwidth while allowing for a machine-based recognition with a high level of accuracy at the final destination.

Without loss of generality and for the sake of clarity, we consider, in what follows, a three-level ROI strategy to decide whether a given block in the frame has to be encoded with a high or low-quality factor or simply discarded. In this study, we have chosen the MJPEG encoder to compress the selected blocks for many reasons: First, MJPEG is more space- and energy-efficient than newer encoders like H.264 and H.265 [68]. These later encoders use motion compensation techniques to reduce temporal redundancy and obtain a better compression ratio, which makes them very computationally expensive [63] [153] [154]. Second, the MJPEG encoder, on the other hand, is considered "solid" because there are no links between frames. If a frame is lost during transmission, the rest of the video will not be compromised, and the error is not propagated to the following frames [155]. This is beneficial in the context of WMSNs, where a hostile channel is presented. Furthermore, the MJPEG can be more required for low

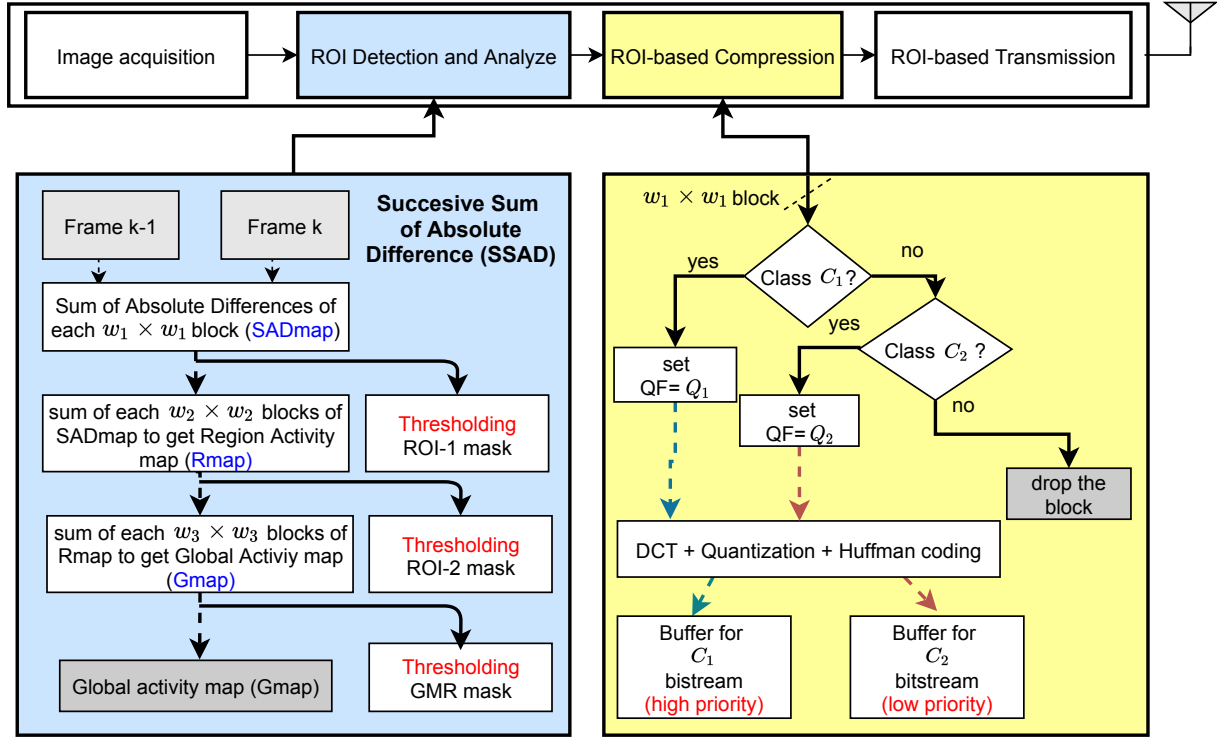


Figure 5.1: The scheme of the proposed strategy for a 3-level ROI-based video coding.

and very low image capture frequencies where temporal redundancy becomes less obvious. This is a typical scenario for a WMSN application where the captured and transmitted images are not necessarily part of a video sequence.

Figure (5.1) illustrates the different steps of the proposed strategy, including our novel ROI detection and classification, which are explained in detail hereafter.

5.2.1 ROI Detection

The proposed ROI detection method aims to classify each frame region based on its activity level. To do so, we introduce the Successive Summation of Absolute Difference (S-SAD) method to compute and classify the activity in the current frame based on a successive summation of different size windowing blocks. First, a SAD between the current frame and a previous frame is calculated on $w_1 \times w_1$ -size non-overlapping blocks :

$$SAD_{map}(x, y) = \frac{1}{w_1^2} \sum_{u=0}^{w_1-1} \sum_{v=0}^{w_1-1} D(w_1x + u, w_1y + v) \quad (5.1)$$

where $x \in 0..M/(w_1 - 1)$ and $y \in 0..N/(w_1 - 1)$ are block coordinates and D is the difference between the current and a previous frame of size $M \times N$. The previous frame is selected based on the activity level and the frame rate of the video. Since a high frame rate creates a low disparity between consecutive frames, it is necessary to select an older frame as a previous frame and vice versa. This step leads to an activity map of w_1^2 times less than the input frame size. The blocks of the activity map that outshine a threshold value are considered to belong to the *ROI-1* which presents the *Local Activity Map*. They are shown in white in Figure (5.2(b)) where Figure (5.2) illustrates the different steps based on a frame sample.

Then, a $w_2 \times w_2$ summation is applied on the obtained SAD_{map} to extract the *Regional Activity Map* R_{map} by substituting w_1 by w_2 in Equation (5.1), that is :

$$R_{map}(x, y) = \frac{1}{w_2^2} \sum_{u=0}^{w_2-1} \sum_{v=0}^{w_2-1} SAD_{map}(w_2x + u, w_2y + v) \quad (5.2)$$

where $x \in 0..M/(w_1w_2 - 1)$ and $y \in 0..N/(w_1w_2 - 1)$. The R_{map} blocks that exceed a threshold value are part of the second region of interest *ROI-2* depicted in Figure (5.2(c)).

Finally and following the same principle, we derive the *Global Activity Map* (G_{map}) of the frame by considering a summation of the *Regional Activity Map* using $w_3 \times w_3$ blocks :

$$G_{map}(x, y) = \frac{1}{w_3^2} \sum_{u=0}^{w_3-1} \sum_{v=0}^{w_3-1} R_{map}(w_3x + u, w_3y + v) \quad (5.3)$$

where $x \in 0..M/(w_1w_2w_3 - 1)$ and $y \in 0..N/(w_1w_2w_3 - 1)$. The thresholding applied on the obtained G_{map} determines the blocks of the third region of interest *ROI-3* (see Figure 5.2(d)). We refer to this final ROI as the global moving region (*GMR*) which should have the property of including the other regions ($ROI-1 \subset ROI-2 \subset ROI-3$) and thus serves as a mask for the previous ROIs to eliminate all blocks initially classified as part of these regions of interest. The excluded blocks are surrounded in red in

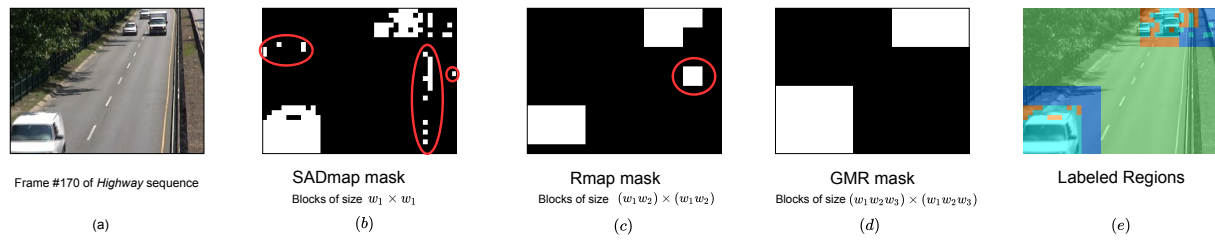


Figure 5.2: ROI construction and GMR mask use for extra blocks elimination.

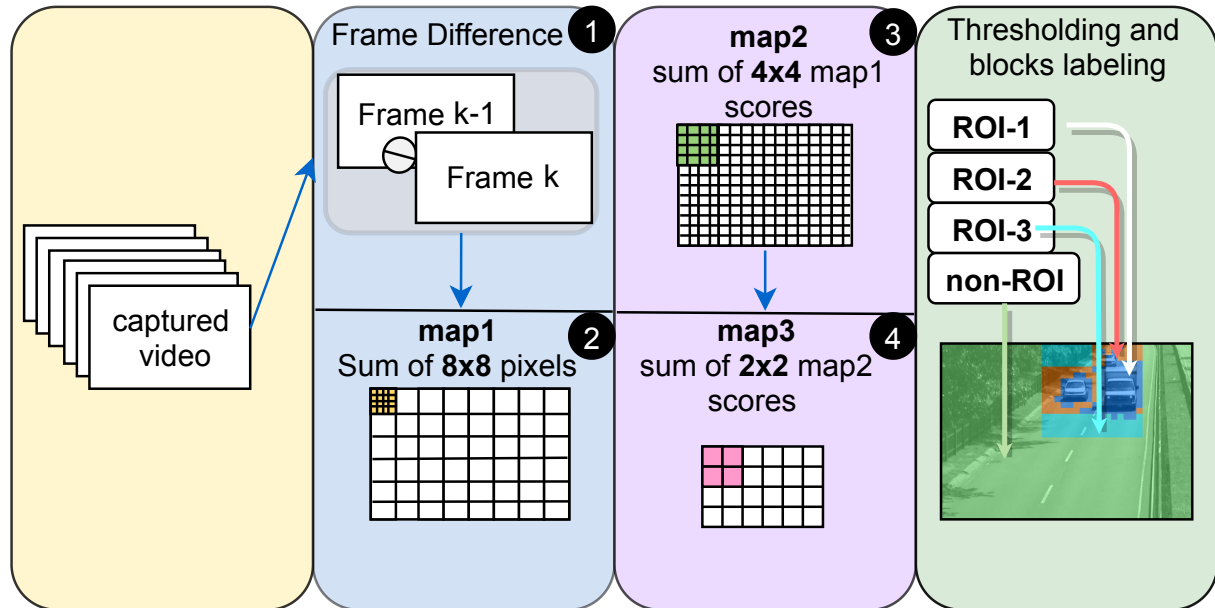


Figure 5.3: The proposed scheme illustration with $w_1 = 8$, $w_2 = 4$ and $w_3 = 2$.

Figure (5.2(b)-(c)). The final obtained ROIs are shown in three different colors in Figure (5.2(e)).

If the frame size is not a multiple of the window size (w_1, w_2 or w_3) for each step, the summation of the elements of the final block is performed only on the remaining pixels. The successive block sizes and the threshold values used in each step have to be appropriately chosen to achieve a good trade-off between computational complexity (large block sizes) and detection precision (small block sizes). According to our numerous tests, we set the block sizes w_1 , w_2 and w_3 , respectively, to 8, 4 and 2, while the threshold values were set empirically following comprehensive testing and experimentation. Figure (5.3) illustrates the proposed scheme based on these latter values.

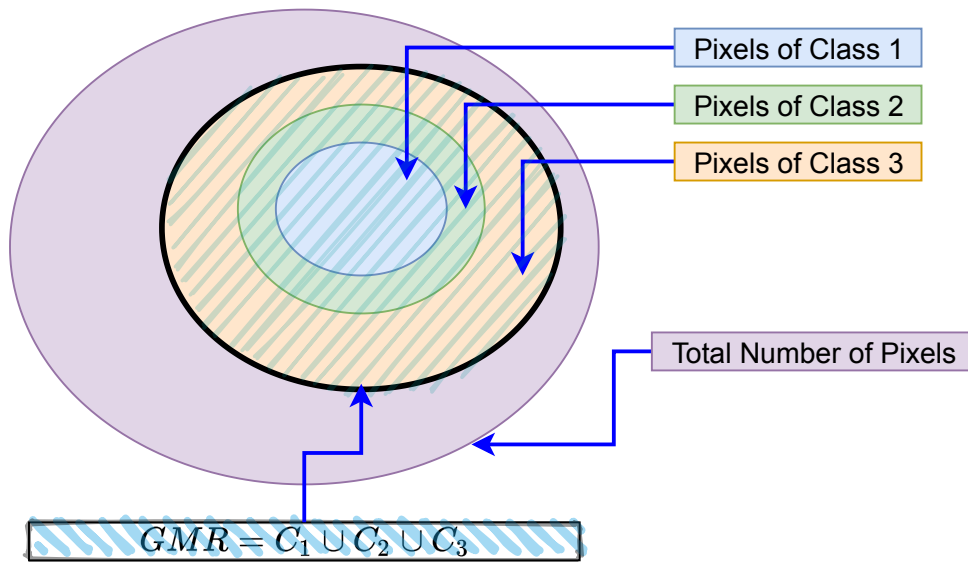


Figure 5.4: Affiliation of each pixel to the classified regions.

5.2.2 ROI Compression and Transmission

First, the initial frame is completely compressed and sent. This first frame is considered a background frame at the destination. For the next frames, only the ROI blocks are taken into account for compression and transmission based on their relative importance by considering the following priority classes shown in Figure (5.4) :

- The first priority class $C_1 = ROI-1$ represents the blocks that are in, and only in the first ROI. Class C_1 blocks having the highest interest are coded with a higher MJPEG quality factor Q_1 before being transmitted ;
- The second priority class $C_2 = ROI-2 - ROI-1$ includes the labeled moving blocks that are in $ROI-2$ but not in $ROI-1$. Class C_2 blocks having a medium interest are coded, before their transmission, with a lower MJPEG quality factor $Q_2 < Q_1$;
- The third priority class $C_3 = GMR - ROI-2$ includes the blocks that are in the GMR but are not in $ROI-2$. These class blocks are considered to be of low interest and are simply dropped.

5.3 Results and Discussion

In order to evaluate the efficiency of our proposed algorithm, we performed extensive simulations and tests using sequences from three standard datasets (Table 5.1), namely ViSOR Dataset [156], Change Detection 2014 Dataset [128] and traffic sequence from MATLAB. The proposed coding scheme is compared to the method in [64] and the standard MJPEG compression where the quality factor QF is set to 50. In our proposed strategy, Q_1 and Q_2 are set to 50 and 20 to encode the blocks of ROI -1 and ROI -2 respectively.

Table 5.1: Used video sequences

Dataset	Video Sequences	Frame Size	fps	# Frames
CDnet	Highway	320×240	30	1700
Dataset	StreetCorner	595×245	25	400
2014 [128]				
ViSOR Dataset	HighwayI	320×240	14	406
[156]	HighwayII	320×240	14	462
	campus	352×288	10	1170
	IntelRoom	320×240	10	300
	Laboratory	320×240	10	880
MATLAB	Traffic	160×120	15	120

5.3.1 Detection Accuracy and Visual Results

Generally, ROI-based techniques suffer from the loss of contextual information. If, additionally, false negatives (i.e. the ROI is falsely determined as non-ROI) increase, the ROI quality degrades significantly. This is why the detection accuracy has to be maximized. Figure (5.5) plots, for the *highway* sequence, the achieved $mTPR$ in estimating the three regions when increasing the distance to the previous frame is considered in the activity map estimation of our scheme. We observe that the older the selected frame, the higher the achieved $mTPR$. We can see that the ROI -3, which includes the two other regions, successfully ranks more than 98% and almost 100% of the movement starting from the second to the last frame for the activity map estimation.

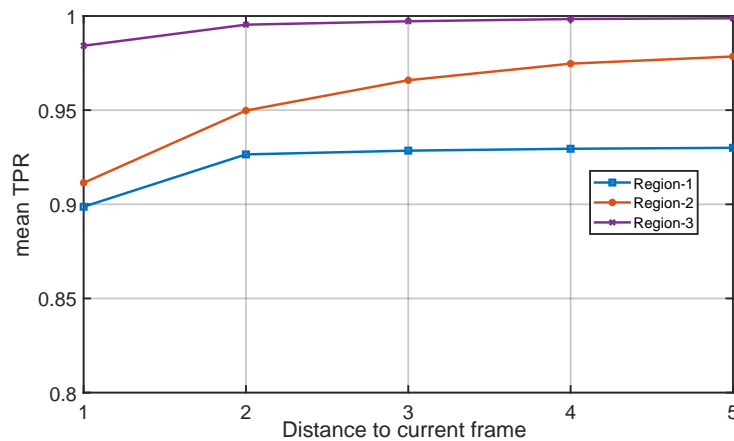


Figure 5.5: Impact of the previous frame selection on TPR for highway sequence

These observations are confirmed by the visual results in Table 5.2 of the reconstructed frames at the reception for the proposed coding method compared to the method in [64] and the standard MJPEG. As can be seen, the whole frame visual quality for our strategy shows better results for all the sequences compared to [64] and slightly lower than those of MJPEG. The detection of the moving region is almost complete and the classification of the regions of interest into high, medium and lower importance is also satisfactory which confirms the high $mTPR$ achieved for $ROI-3$ as well as for $ROI-1$ and $ROI-2$ in Figure (5.5).

Furthermore, based on our experiments, our strategy does not suffer from the error propagation problem encountered by the approach in [64]. This is due to the fact that our method selects larger regions in which the movement occurred before determining the ROI to be compressed and transmitted, which significantly lowers the probability of a region selection error (loss of context). Our strategy ensures effective detection for different scenarios, both indoor as is the case for *laboratory* and *intelligentroom* sequences and outdoor as can be seen in the other sequences. The detection accuracy is not affected by dark background scenarios, like in the case of *StreetCornerAtNight* sequence as illustrated in the samples in the ROI column in Table 5.2. It is also worth noting that the speed of the movement does not affect the detection performance as observed in the results of the *laboratory* sequence. This is one key advantage of the

adopted frame difference-based detection technique.

5.3.2 Quantitative Results: Image Quality

We first evaluate the quality of the images using three metrics, namely PSNR, SSIM and VIF, to estimate the difference between the original images and the reconstructed images. This provides an estimate of the information loss resulting from the compression phase on the different considered video sequences. Table 5.3 shows the mean PSNR, SSIM and VIF values for each sequence. The bold values represent the best results, while the underlined ones represent the second-best results. We note that MJPEG performs the best in terms of the PSNR and VIF for all sequences. This was to be expected since the entire image is transmitted after being encoded with a high-quality factor. Our proposed method gives the second-best results for all metrics and outperforms MJPEG in terms of the SSIM measure for specific sequences (*Intellegentroom* and *Highway*). This superiority is directly due to the nature of the scene background, which is distinguished by significant stability (no change over time). Thus, the high quality of the background blocks is maintained over time. However, we observe a considerable decrease in the quality of the reconstructed frames for some sequences (such as *HighwayI* and *HighwayII*). This degradation is due to the low frame rate of the video with the significant motion that occurs in the scene. Moreover, the large size of the moving objects may cause inaccurate ROI detection, resulting in a substantial loss of contextual information. Furthermore, a high movement leads to more ROI-2 zones coded with low quality ($Q_2 = 20$), which decreases the frame quality.

For a more refined analysis, we plotted the curves representing the three metrics on a per-frame basis for each considered video sequence in Figures (5.6), (5.7) and (5.8). It is clearly shown that the proposed strategy guarantees an acceptable high quality of the received frames under a very low size of the data to be transmitted as will be shown in Section 5.3.3. The method registers about 30 dB PSNR or higher for all the sequences

Table 5.2: Visual binary mask for the moving region.







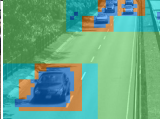




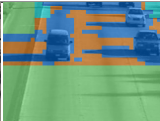




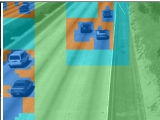



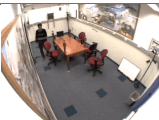










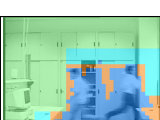








Seq.	Original	ROI	Proposed	MJPEG	[64]
Traffic #10					
SSIM	-	-	0.9586	0.8992	0.8915
Highway #687					
SSIM	-	-	0.8450	0.9313	0.8364
HighwayI #216					
SSIM	-	-	0.9135	0.9613	0.8755
HighwayII #484					
SSIM	-	-	0.8521	0.9110	0.8397
IntellegentRoom #231					
SSIM	-	-	0.9287	0.9428	0.9129
Campus #114					
SSIM	-	-	0.8783	0.9408	0.8374
Laboratory #806					
SSIM	-	-	0.9195	0.9476	0.9023
StreetCorner #884					
SSIM	-	-	0.9311	0.9891	0.9163

Table 5.3: Overall mean quality metrics

Sequence	proposed			ref. [64]			MJPEG		
	PSNR	SSIM	VIF	PSNR	SSIM	VIF	PSNR	SSIM	VIF
Highway	<u>31.7414</u>	0.7865	<u>0.6042</u>	30.4667	0.7053	0.5385	32.7808	<u>0.7700</u>	0.7351
HighwayI	28.8923	<u>0.6716</u>	<u>0.5744</u>	<u>31.8053</u>	0.6138	0.4874	37.9583	0.8374	0.7934
HighwayII	28.4600	<u>0.7055</u>	<u>0.4637</u>	<u>29.3400</u>	0.6965	0.4506	33.0100	0.8208	0.7207
campus	<u>31.3614</u>	<u>0.7055</u>	<u>0.4637</u>	29.5200	0.6965	0.4501	35.7400	0.8208	0.7207
intellegentroom	<u>31.7727</u>	0.8036	<u>0.5916</u>	30.4667	0.7053	0.5385	32.7808	<u>0.7700</u>	0.7351
laboratory	<u>32.1748</u>	<u>0.6214</u>	<u>0.5748</u>	30.9583	0.5894	0.5297	34.6275	0.6790	0.7492
Traffic	<u>30.0569</u>	<u>0.6559</u>	<u>0.6230</u>	28.2246	0.5710	0.5030	30.4093	0.6625	0.6493
StreetCornerAtNight	<u>33.3932</u>	<u>0.9181</u>	<u>0.9209</u>	32.1294	0.8815	0.8724	42.7939	0.9690	0.9514

as depicted in Figure (5.6). Our strategy shows at most 4 dB lower PSNR values when compared to MJPEG and a higher PSNR, for all the sequences, with respect to the method in [64].

For the SSIM values, the proposed strategy registers values varying from 0.6 to 1 as shown in Figure (5.7). Most of the first received frames exhibit high SSIM values (above 0.95) which degrade over time. Deterioration is due to the increasing amount of blocks being classified as *ROI-2*, compressed with lower quality ($Q_2 = 20$). Nevertheless, we still obtain higher SSIM values with respect to the method proposed in [64]. Similar results are obtained based on the VIF metric as shown in Figure (5.8). The registered values ranging almost from 0.4 to 1 for the different considered sequences. Some sequences register high degradation in terms of VIF like in the *campus* clip that undergoes a degradation as high as 0.4. This is due to the complexity of the corresponding background. The same degradation is noticed in [64] but with slightly better results for our strategy.

Overall, the proposed method outperforms the method in [64]. We can say that the proposed strategy is able to ensure acceptable quantitative results by the transmission of only a few data which makes it a good alternative to the MJPEG coding method where the transmission of the whole frame is needed.

No-reference Image Quality assessment Figure (5.9) depicts the obtained BRISQUE score. We note that the original frames account for the best performances with the BRISQUE metric recording the lowest values. MJPEG frames encoded with a $Q_1 = 50$ presents also good performances which are comparable to the original frames. For the proposed strategy, it is noticed that the performances are slightly related to the type of the sequence. Results, in almost all sequences, are better than those in [64]. BRISQUE scores of the proposed strategy are also comparable to the original and MJPEG for some sequences such as *traffic*, *intelligentroom*, and *laboratory*. The results show that coding artifacts affects negatively the BRISQUE score. Another assumption is that the type of the environment and background may affect the perceptual quality as analyzed in our case (better BRISQUE score for indoor environments). The noisy background is also a reason for the high BRISQUE score since the proposed strategy and the method in [64] consider sending only the ROI without an update of the background information. Through the analysis done in [114], the oscillatory aspect of the quality measure is mainly due to the inability of these metrics to quantify the quality level of the videos, which is considered a disadvantage of the BRISQUE metric as a measure for smart surveillance systems.

5.3.3 Bitrate Gain

Reducing the bitrate saves a colossal amount of transmission energy and network bandwidth which consequently avoids or at least limits congestion situations and allows better delivery conditions. These benefits are the main focus of the proposed

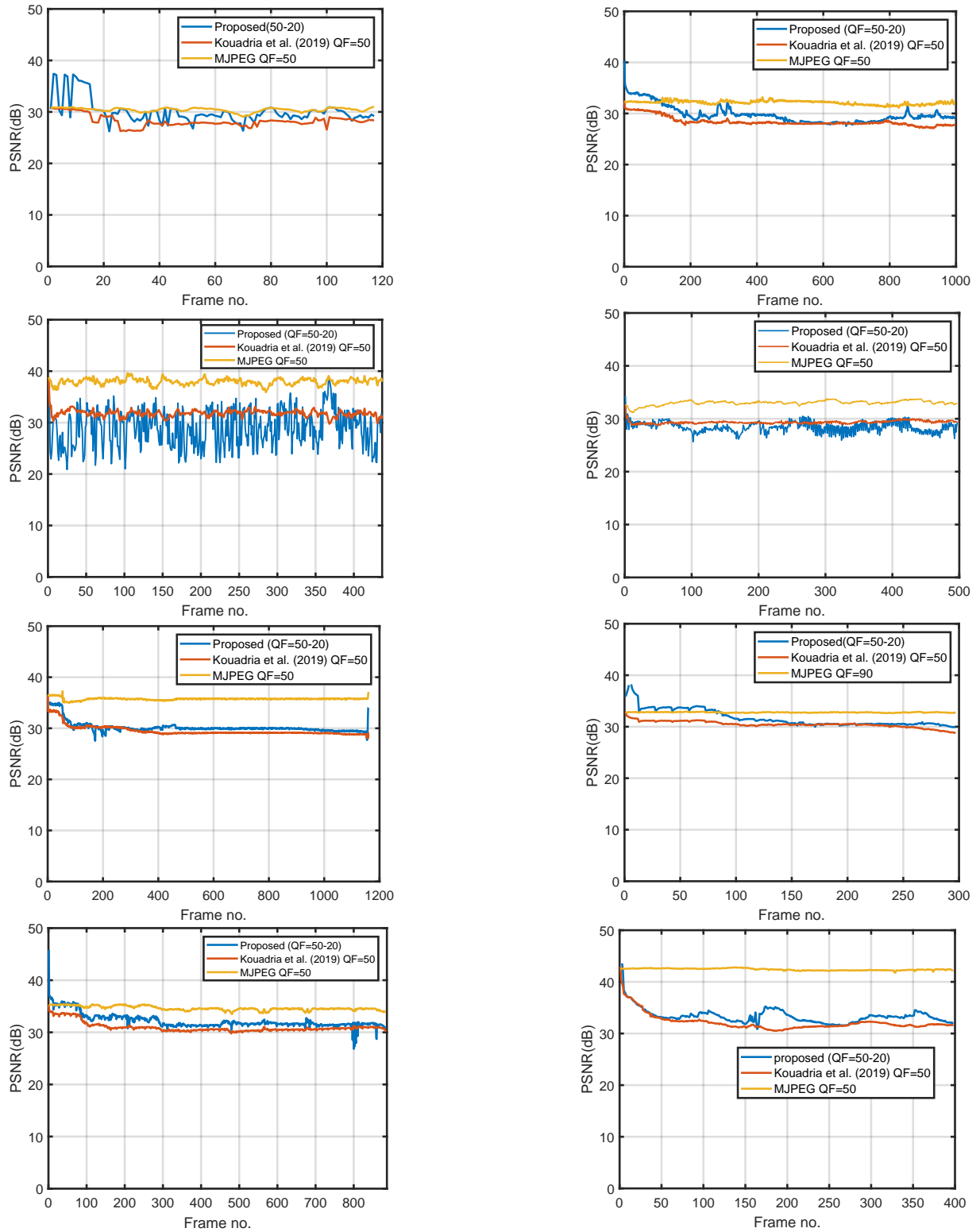


Figure 5.6: PSNR results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.

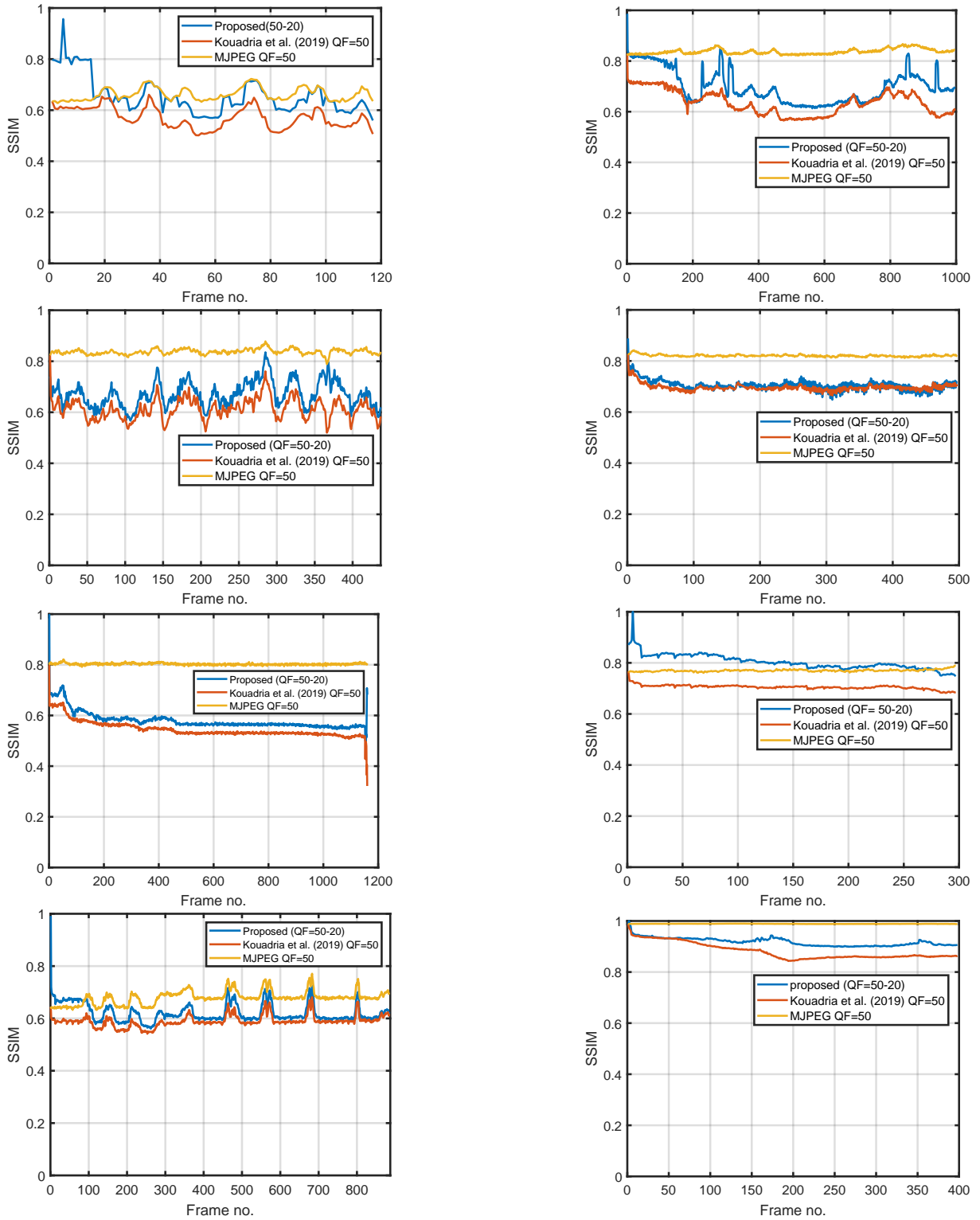


Figure 5.7: SSIM results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.

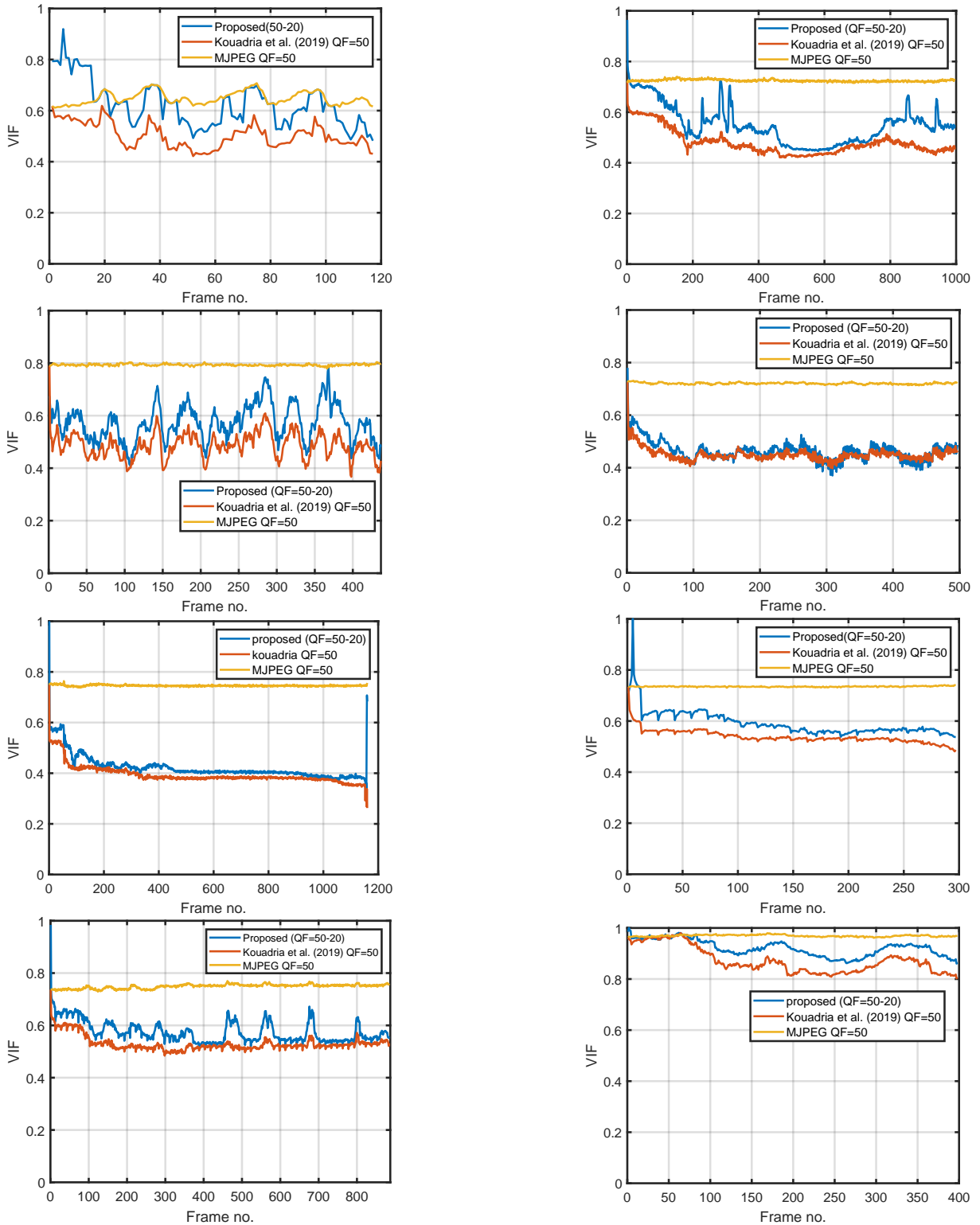


Figure 5.8: VIF results for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight.

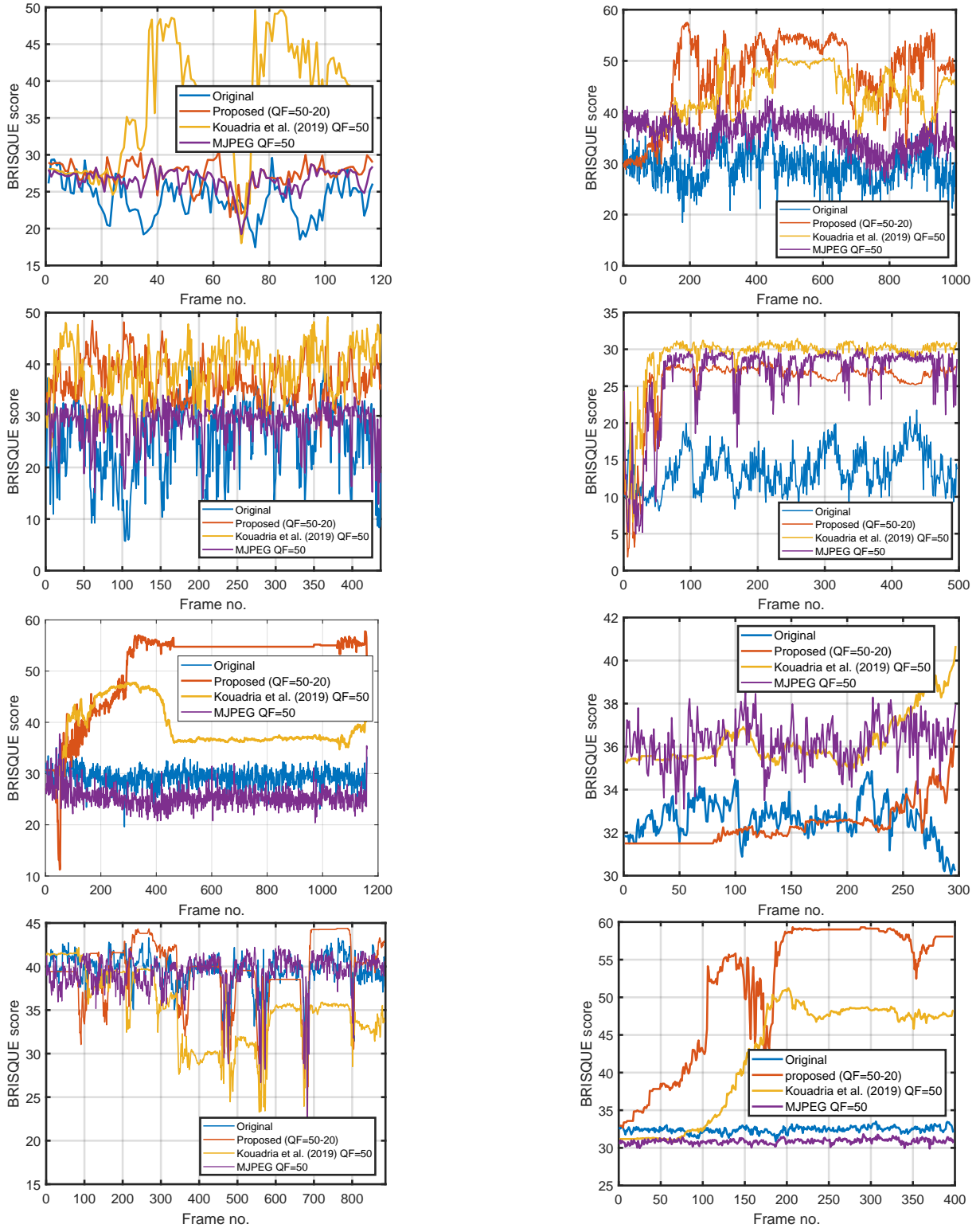


Figure 5.9: BRISQUE scores for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight

contribution. In what follows, we reason in terms of the amount of data to be delivered after the compression of each of the frames composing a given video sequence. This quantity of data gives an idea of the bitrate needed for the application's desired capture rate. Figure (5.10) shows the amount of data to be transmitted on a per-frame basis when adopting our proposed method, MJPEG and [64]. We can see that the required bitrate for the suggested strategy is slightly higher than in [64] for most of the sequences. This was expected since the size of ROI in our case is larger but in turn, our strategy ensures the delivery of a higher quality ROI at the expense of a slightly higher bitrate. For instance, with the same quality level, our method requires a mean bitrate of 3.358 kB/s for the *campus* sequence ($fps = 10$), which is 27 times less than the required bitrate when adopting MJPEG (93.06 kB/s) which represents a saving of 96.4%. For the *highway* sequence ($fps = 25$), we achieve a saving of about 76.3% of the required bitrate since it drops from 263.25 kB/s to 62.65 kB/s.

5.3.4 The Impact of Quality Degradation on Object Recognition

We use the real-time object detection system YOLOv3 [157] as a machine-based monitoring system on the received frames to extract and recognize moving objects. The lightweight YOLOv3 network architecture (Figure (5.11)) contains 13 convolution layers and 6 max-pooling layers. The used tiny-YOLOv3 is trained on the COCO dataset [158] to classify objects into 80 classes. Due to its competitive accuracy and speed, and its robustness in detecting different types of objects, YOLOv3 algorithms have been widely applied in industries, such as manufacturing and the military. We aim to assess the efficiency of the proposed strategy in ensuring a high recognition accuracy under a very low bitrate. The YOLOv3 model is used as a "black box" meaning that no study on the performance of the model on the used dataset is performed. We take it as it is. We compare the recognition accuracy of the proposed strategy to the original non-compressed frames, MJPEG compressed frames, and compressed frames using [64].

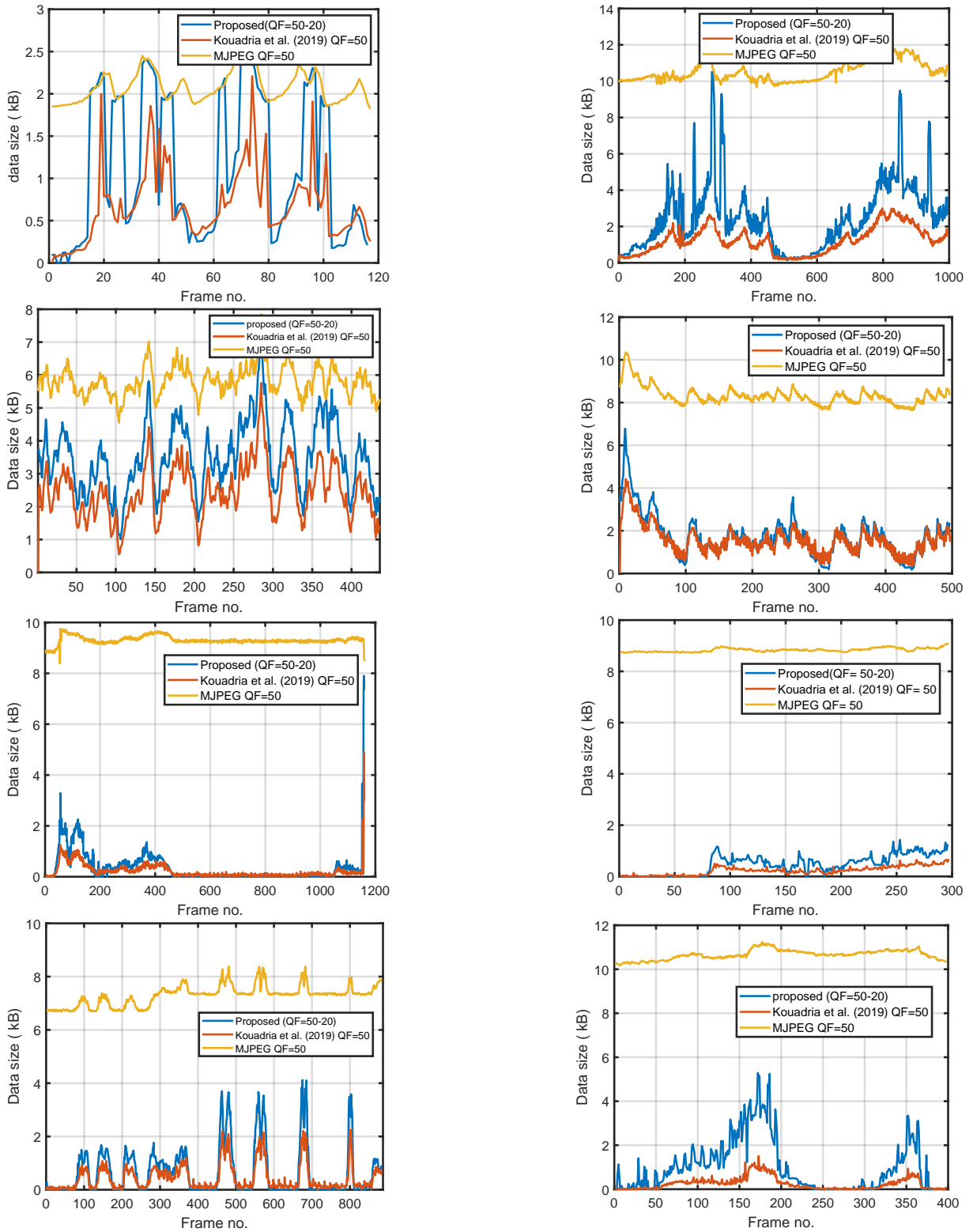








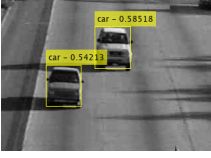
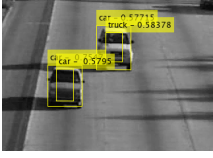
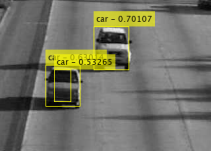
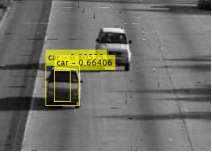






















Figure 5.10: Amount of data to transmit per frame for different sequences (left to right top-bottom scanning): Traffic, Highway, HighwayI, HighwayII, Campus, Intellegentroom, Laboratory, StreetCornerAtNight

Table 5.4: Bounding box insertion results for the used dataset.

Sequence	Proposed	Original	MJPEG	[64]
Traffic #88				
Highway #795				
HighwayI #126				
HighwayII #268				
Campus #71				
IntelligentRoom #252				
Laboratory #686				
StreetCornerAtNight #884				

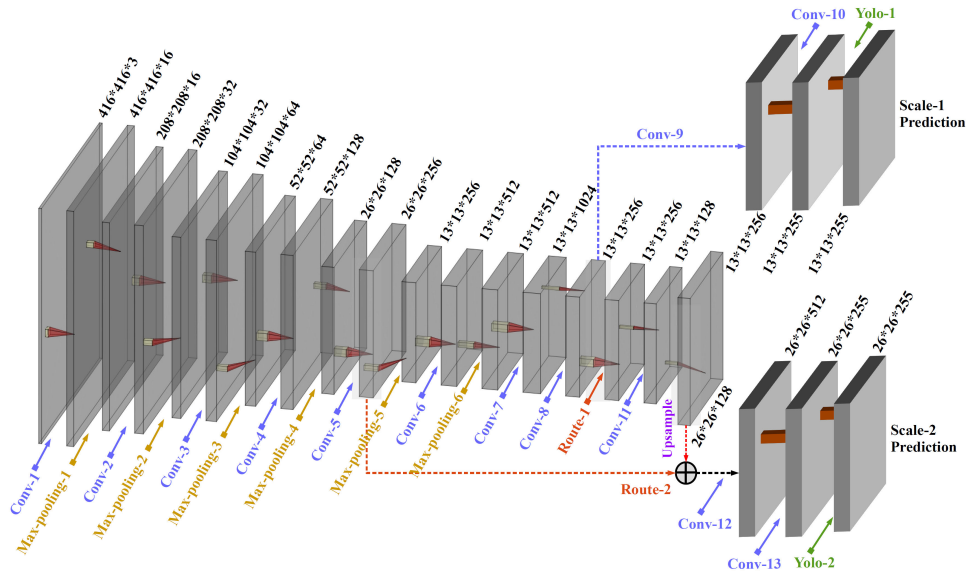


Figure 5.11: Tiny-YOLOv3 network Architecture (redrawn from [151])

Table 5.4 shows the recognition results for sample frames from the used datasets and demonstrates the competitiveness of our method to enhance the recognition accuracy at the destination. For all sequences, we achieve higher recognition accuracy compared to [64]. The results are still comparable to the original and MJPEG frames or even better in many cases compared to MJPEG. At first impression, results show that the compression quality degradation hurts the recognition accuracy. This conclusion is supported by the overall recognition results depicted in Figure (5.12) and Figure (5.13) which represents the performance of the recognition process, for all the dataset. Figure (5.12) plots the number of the detected objects and shows that unsurprisingly, original frames achieve the highest score for all the sequences. Our proposed strategy, overall, shows the second-best results. Preserving only a high-quality compression of the ROI while ensuring a good ROI detection is sufficient to enable more accurate smart tasks at the destination like object recognition. As for the recognition accuracy, Figure (5.13) shows the superiority of our method which allows smart machine-based tasks at a significantly low bitrate and energy budgets as will be shown in the following section.

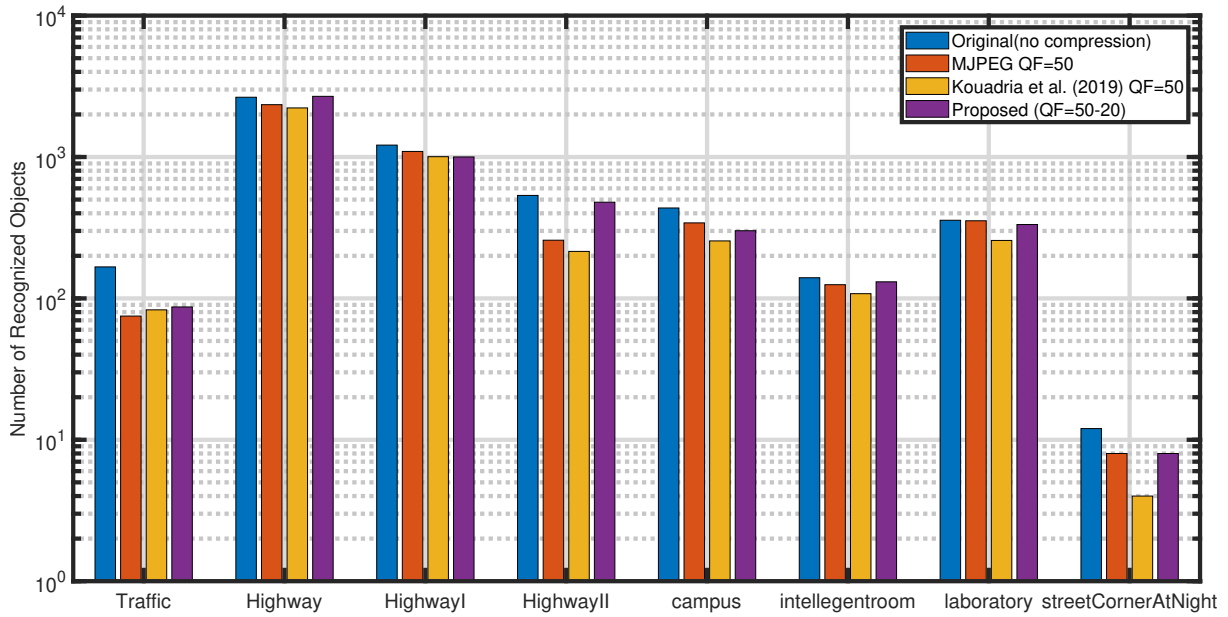


Figure 5.12: Performance of recognition: Number of detected objects

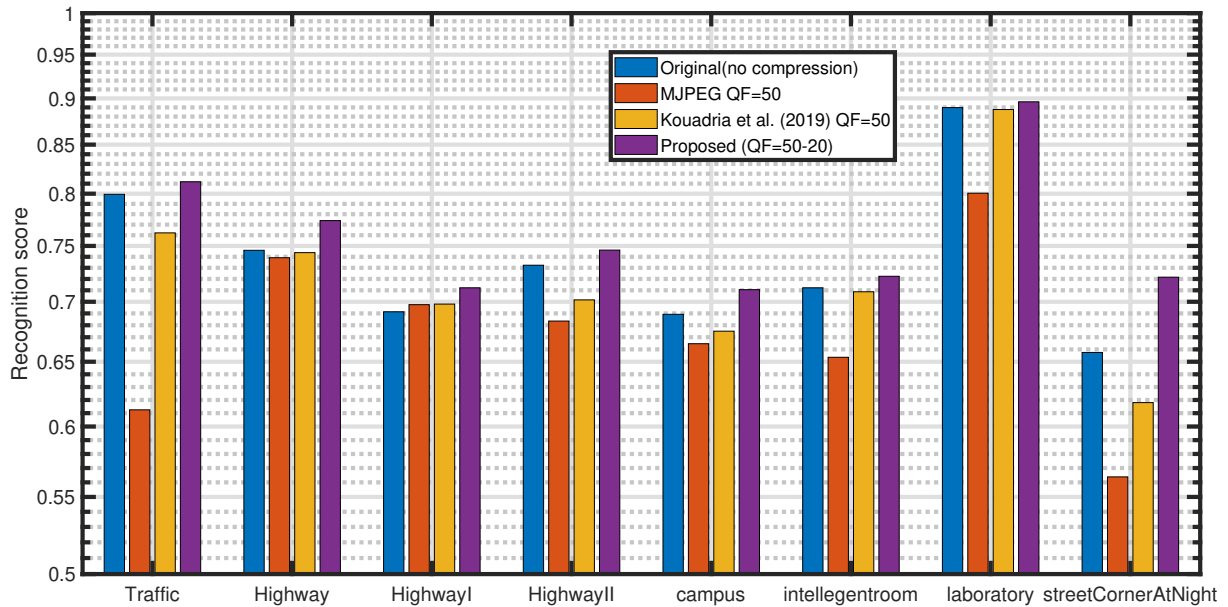


Figure 5.13: Performance of recognition: Recognition accuracy

5.3.5 Computational Complexity and Energy Consumption

The advantage of the proposed detection method is its very low complexity since the summation is the most used operation in all the steps. Just a very low number of divisions is used to normalize the scores in the activity maps. We show through this section the effectiveness of the proposed strategy in terms of computational complexity. Let $E_{processing}$ be the energy consumption by a visual sensor to detect the ROI and compress the corresponding data, then we can write :

$$E_{processing} = E_{detection} + E_{compression} \quad (5.4)$$

where $E_{compression}$ is the JPEG-like compression energy cost and the $E_{detection}$ is the required energy to detect the ROI and classify them. This latter can be estimated based on the number of operations needed by each step of the proposed strategy which we provide in Table 5.5. That is :

$$E_{detection} = E_{SADmap} + E_{Rmap} + E_{Gmap} + E_{Thresh} \quad (5.5)$$

For a given step $s = SAD_{map}, R_{map}, G_{map}, Thresh$, the necessary energy (E_s) can be estimated using the following formula :

$$E_s = \sum_{op} N_{s,op} \times \varepsilon_{op} \times Cycles_{op} \quad (5.6)$$

Where $N_{s,op}$ is the required number of operations op (addition, subtraction, division or thresholding) to perform the step, ε_{op} is the energy consumption of operation op and $Cycles_{op}$ is the number of required cycles to execute operation op . These two latter parameters depend on the underlying processor. In embedded micro-controllers, we know that the energy cost of the addition, subtraction, threshold and absolute operations are the same : $\varepsilon_{abs} = \varepsilon_{sub} = \varepsilon_{thresh} = \varepsilon_{add}$ and $Cycles_{abs} = Cycles_{sub} =$

$Cycles_{thresh} = Cycles_{add}$. Moreover, knowing that $w_2 = w_1/2$ and $w_3 = w_1/4$, we get to the following formula :

$$E_{detection} = NM \left(1 + \frac{4}{w_1^2} + \frac{4}{w_1^4} + \frac{64}{w_1^6} \right) \varepsilon_{add} Cycles_{add} + NM \left(\frac{1}{w_1^2} + \frac{4}{w_1^4} + \frac{64}{w_1^6} \right) \varepsilon_{div} Cycles_{div} \quad (5.7)$$

To estimate $E_{compression}$, we have to count for the MJPEG compression cost for each block including the DCT, the quantization and the Huffman coding costs. We adopt the model provided in [150] where three implementations of the float IJG, slow IJG and fast IJG JPEG-based compression were discussed. It has been shown that the slow IJG [159] achieves the same quality performances as float IJG while the fast IJG [160] has a significant quality loss. In this work, we have chosen the slow IJG implementation which has a compression energy cost of 192.28 μ J per 8×8 block.

Table 5.5: Computational Cost of each step

Step	Operation	Window	Complexity (# of operations)
SAD_{map}	Add	w_1	$NM(w_1^2 - 1)/w_1^2$
	Sub		NM/w_1^2
	Abs		NM/w_1^2
	Div		NM/w_1^2
R_{map}	Add	$w_2 = w_1/2$	$4NM/w_1^4$
	Div		
G_{map}	Add	$w_3 = w_1/4$	$64NM/w_1^6$
	Div		
Thresh.	SAD_{map}	w_1	NM/w_1^2
	R_{map}	$w_2 = w_1/2$	
	G_{map}	$w_3 = w_1/4$	

Energy Consumption Discussion First, the added cost of the proposed ROI detection step is figured out. Table 5.6 records the mean consumed energy per frame for the detection phase ($E_{detection}$) versus the overall processing energy ($E_{processing}$) for three video sequences of different frame sizes (*campus*, *highway* and *traffic*). It shows that the

extra consumed energy depends on the size of the images and remains moderate as it does not exceed 7% of the processing energy. Table 5.7 gives a better insight into the processing energy of our proposal by indicating the variation of this consumption around the mean value to be compared with that of MJPEG. We can see that our strategy allows a significant energy saving which can be as high as 93% for the *campus* sequence for instance and could achieve 90% or more depending on the amount of activity during the surveillance task for the two other sequences. We note a lower energy saving of about 50% for the *traffic* sequence due mainly to the small size of its frames making the chosen values for w_1 , w_2 , and w_3 less adequate. To save even more energy, these values must be more efficiently adjusted to the frame size. The economy also depends on the size of the moving objects in the scene. Small objects imply small ROIs and thus fewer data to compress and vice versa. We observe high deviation values (close to or exceeding the mean value) which reflect a significant variance in energy expenditure. This is attributable to a varying activity level in the different parts of the sequence.

Frames that exhibit a high activity map require much energy consumption, while low activity periods lead to very limited energy usage. The *campus* sequence, for instance, records energy values that range from 0.9 mJ for extremely little activity and more than 211 mJ for a frame (#1164) that even exceeds the MJPEG energy. This is the consequence of three successive outlier frames that are completely noisy. In the first outlier frame #1164, all the blocks are considered for coding resulting in high energy consumption. While for the subsequent frames, only some blocks are considered for coding due to the low detected difference. This outliers problem is a common problem that has forced the system to react as in the basic approach where all the frame blocks were considered as being part of the ROI [161].

Figure (5.14) shows the evolution of the consumed energy per frame, for three sequences with different frame sizes, as well as the achieved quality in terms of PSNR. The energy curves confirm our earlier findings. The oscillation in energy consumption

Table 5.6: Per frame energy cost (mJ) of our ROI detection

Sequence	$E_{detection}$	$E_{processing}$	% extra cost
<i>campus</i>	0.8827	14.24	6.20%
<i>highway</i>	0.6699	16.02	4.18%
<i>traffic</i>	0.1671	20.22	0.83%

Table 5.7: Per-frame energy consumption (mJ).

Sequence	Proposed				MJPEG mean	saving (%) w/r MJPEG
	max	min	std. dev.	mean		
<i>campus</i>	211.14	0.90	24.93	14.24	205.92	93.08
<i>highway</i>	53.38	0.90	10.96	16.02	156	89.74
<i>traffic</i>	40.18	0.23	16.11	20.22	39	48.16

is directly related to the size of the ROI. It is illustrated that the method yields lower energy consumption than the classical (MJPEG) compression method that registers a sufficiently stable higher value. We note that the limited amount of data to be processed and sent, allowing for a significant reduction in energy consumption, does not impact the quality of the video, which remains rather stable in the range of 30..35 dB.

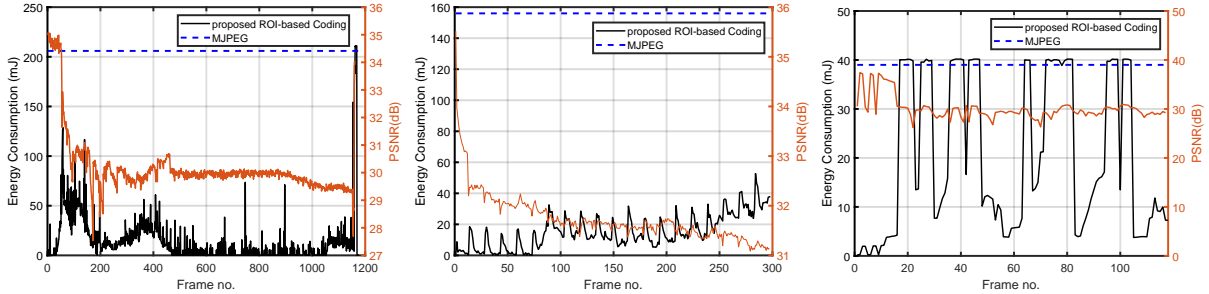


Figure 5.14: Total processing energy consumption and the corresponding PSNR for sequences with different frame sizes. *campus* 352×288 , *highway* 320×240 , *traffic* 160×120

The method could show further gains when it is studied in a network-based scenario, considering a significant number of wireless sensor nodes to cover a specific zone for intelligent surveillance tasks. It can further reduce network congestion and multiply the gains in the network's resources. It can also operate in a contextual paradigm where it is not always necessary to send the detected movement into a specifically covered zone [162].

5.3.6 Conclusion

We proposed an ROI-based video coding strategy for human-based and machine-based video surveillance monitoring using WMSN. The method exhibits a low bitrate compared to conventional MJPEG through the proposed low-cost ROI detection method. The strategy also aims to reduce significantly processing energy consumption in the sensor node. Through the evaluation of different sequences, it has been shown that the energy savings could reach 90% with a slight sacrifice in the quantitative and perceptual quality of the non-ROI. It has also been shown that the quality sacrificed of the non-ROI does not influence the intelligent tasks at the destination but enhances them by virtue of the content-aware strategy used. The proposed video coding strategy could be adopted for large-scale video monitoring in an edge-cloud processing paradigm using WMSN, where in-network-based scenarios should be elaborated and assessed. Further study is recommended to illustrate unusual coding conditions and issues like the occurrence of outlier frames and/or outlier blocks during the processing and transmission of the frame.

General conclusion and perspectives

Wireless Surveillance systems are poised to play a crucial role in the future of communication systems, as evidenced by the rapid advancements in the field over the last decade. Further progress in the development of efficient visual sensor nodes will solidify the substantial progress made in recent years, and pave the way for new and innovative applications of wireless surveillance technology.

In conclusion, this thesis presents a methodical examination of the role of wireless surveillance systems using WMSN in modern communication networks. Through a comprehensive literature review and the development of novel optimization approaches, this work highlights the impact of low energy and bitrate on the quality of service in WVS. The proposed ROI-based video coding algorithms, such as the BIRD algorithm, offer a promising solution by effectively adapting to the resource constraints of WMSN and achieving high accuracy in classifying ROIs, as well as reducing the bitrate required for data transmission. Our results shed light on the potential of ROI-based video coding to enable advanced video content analysis tasks, such as facial recognition and object tracking, as demonstrated by the 22% improvement using YOLOv3 as an inference model.

These findings open up exciting avenues for further research and development in this field, such as exploring some of the points:

- Implementing fast transformation algorithms in the image compression step.

- The development of adapted MAC protocols that meet content-aware requirements.
- Further testing and implementation of these techniques in real-case platforms like SenseVid will pave the way for new concepts of transmitting multimedia data in WMSN with the lowest possible resources.
- Exploring the ROI-based video coding with the new codecs like H264 and HEVC may enable their use in WMSN contrary to the actual case of new codecs and their non-adaptability for WMSN.
- Exploring the VCM approach in a wider way is an excellent choice for more adaptability between ROI-based video coding and Machine dedicated services and systems.

This thesis demonstrates the significance of ROI-based video coding in wireless surveillance systems using WMSN and the potential for continued exploration and innovation in this field.

Bibliography

- [1] Omar Elharrouss, Noor Almaadeed, and Somaya Al-Maadeed. A review of video surveillance systems. *Journal of Visual Communication and Image Representation*, 77:103116, 2021.
- [2] Gopal Ghosh, Monica Sood, Sahil Verma, et al. Internet of things based video surveillance systems for security applications. *Journal of Computational and Theoretical Nanoscience*, 17(6):2582–2588, 2020.
- [3] Gyula Simon, Miklós Maróti, Ákos Lédeczi, György Balogh, Branislav Kusy, András Nádas, Gábor Pap, János Sallai, and Ken Frampton. Sensor network-based countersniper system. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 1–12, 2004.
- [4] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal. Analysis of a prediction-based mobility adaptive tracking algorithm. In *2nd International Conference on Broadband Networks, 2005.*, pages 753–760. IEEE, 2005.
- [5] Mauricio Castillo-Effer, Daniel H Quintela, Wilfrido Moreno, Ramiro Jordan, and Wayne Westhoff. Wireless sensor networks for flash-flood alerting. In *Proceedings of the Fifth IEEE International Caracas Conference on Devices, Circuits and Systems, 2004.*, volume 1, pages 142–146. IEEE, 2004.

- [6] Konrad Lorincz, David J Malan, Thaddeus RF Fulford-Jones, Alan Nawoj, Antony Clavel, Victor Shnayder, Geoffrey Mainland, Matt Welsh, and Steve Moulton. Sensor networks for emergency response: challenges and opportunities. *IEEE pervasive Computing*, 3(4):16–23, 2004.
- [7] Tia Gao, Dan Greenspan, Matt Welsh, Radford R Juang, and Alex Alm. Vital signs monitoring and patient tracking over a wireless network. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 102–105. IEEE, 2006.
- [8] Geoffrey Werner-Allen, Konrad Lorincz, Mario Ruiz, Omar Marcillo, Jeff Johnson, Jonathan Lees, and Matt Welsh. Deploying a wireless sensor network on an active volcano. *IEEE internet computing*, 10(2):18–25, 2006.
- [9] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal. Wireless sensor network survey. *Computer networks*, 52(12):2292–2330, 2008.
- [10] Mariam Al Nuaimi, Farag Sallabi, and Khaled Shuaib. A survey of wireless multimedia sensor networks challenges and solutions. In *2011 International Conference on Innovations in Information Technology*, pages 191–196. IEEE, 2011.
- [11] Xiaoding Wang, Sahil Garg, Hui Lin, Jia Hu, Georges Kaddoum, Md Jalil Piran, and M Shamim Hossain. Toward accurate anomaly detection in industrial internet of things using hierarchical federated learning. *IEEE Internet of Things Journal*, 9(10):7110–7119, 2021.
- [12] Ali Nauman, Yazdan Ahmad Qadri, Muhammad Amjad, Yousaf Bin Zikria, Muhammad Khalil Afzal, and Sung Won Kim. Multimedia internet of things: A comprehensive survey. *IEEE Access*, 8:8202–8250, 2020.
- [13] Amira Boulmaiz, Nouredine Doghmane, Saliha Harize, Nasreddine Kouadria, and Djemil Messadeg. The use of WSN (wireless sensor network) in the surveil-

- lance of endangered bird species. In *Advances in ubiquitous computing*, pages 261–306. Elsevier, 2020.
- [14] James Pierce. Smart home security cameras and shifting lines of creepiness: A design-led inquiry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.
- [15] Tan Zhang, Aakanksha Chowdhery, Paramvir Bahl, Kyle Jamieson, and Suman Banerjee. The design and implementation of a wireless video surveillance system. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 426–438, 2015.
- [16] Yun Ye, Song Ci, Aggelos K Katsaggelos, Yanwei Liu, and Yi Qian. Wireless video surveillance: A survey. *IEEE Access*, 1:646–660, 2013.
- [17] S Muthukarpagam, V Niveditta, and S Neduncheliyan. Design issues, topology issues, quality of service support for wireless sensor networks: Survey and research challenges. *International Journal of Computer Applications*, 1(6):1–4, 2010.
- [18] Tifenn Rault, Abdelmadjid Bouabdallah, and Yacine Challal. Energy efficiency in wireless sensor networks: A top-down survey. *Computer networks*, 67:104–122, 2014.
- [19] Da Li, Zhang Zhang, Kai Yu, Kaiqi Huang, and Tieniu Tan. Isee: an intelligent scene exploration and evaluation platform for large-scale visual surveillance. *IEEE Transactions on Parallel and Distributed Systems*, 30(12):2743–2758, 2019.
- [20] Jianguo Chen, Kenli Li, Qingying Deng, Keqin Li, and S Yu Philip. Distributed deep learning model for intelligent video surveillance systems with edge computing. *IEEE Transactions on Industrial Informatics*, 2019.
- [21] Tong Sun, Yongquan Xia, and Yong Gan. Discussion on integration of urban video surveillance system. *Procedia Engineering*, 15:3255–3259, 2011.

- [22] Yuan He, Junchen Guo, and Xiaolong Zheng. From surveillance to digital twin: Challenges and recent advances of signal processing for industrial internet of things. *IEEE Signal Processing Magazine*, 35(5):120–129, 2018.
- [23] Naor Kalbo, Yisroel Mirsky, Asaf Shabtai, and Yuval Elovici. The security of ip-based video surveillance systems. *Sensors*, 20(17):4806, 2020.
- [24] Li-minn Ang, Kah Phooi Seng, Li Wern Chew, Lee Seng Yeong, and Wai Chong Chia. *Wireless multimedia sensor networks on reconfigurable hardware*. Springer, 2013.
- [25] Karan Nair, Janhavi Kulkarni, Mansi Warde, Zalak Dave, Vedashree Rawalgaonkar, Ganesh Gore, and Jonathan Joshi. Optimizing power consumption in iot based wireless sensor networks using bluetooth low energy. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 589–593, 2015.
- [26] Wenjing Guo, Cairong Yan, and Ting Lu. Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing. *International Journal of Distributed Sensor Networks*, 15(2):1550147719833541, 2019.
- [27] Sultan Mahmood Chowdhury and Ashraf Hossain. Different energy saving schemes in wireless sensor networks: A survey. *Wireless Personal Communications*, 114(3):2043–2062, 2020.
- [28] Dinesh Kumar Sah and Tarachand Amgoth. Parametric survey on cross-layer designs for wireless sensor networks. *Computer Science Review*, 27:112–134, 2018.
- [29] Aliaa A A Youssif, Atef Zaki Ghalwash, and Others. Energy aware and adaptive cross layer scheme for video transmission over wireless sensor networks. *IEEE Sensors Journal*, 16(21):7792–7802, 2016.

- [30] Moufida Maimour. Maximally radio-disjoint multipath routing for wireless multimedia sensor networks. In *Proceedings of the 4th ACM workshop on Wireless multimedia networking and performance modeling*, pages 26–31, 2008.
- [31] Moufida Maimour, Houda Zeghilet, and Francis Lepage. Cluster-based routing protocols for energy-efficiency in wireless sensor networks. *Sustainable Wireless Sensor Networks*, pages 167–188, 2010.
- [32] M Aykut Yigitel, Ozlem Durmaz Incel, and Cem Ersoy. Design and implementation of a qos-aware mac protocol for wireless multimedia sensor networks. *Computer Communications*, 34(16):1991–2001, 2011.
- [33] Navrati Saxena, Abhishek Roy, and Jitae Shin. Dynamic duty cycle and adaptive contention window based qos-mac protocol for wireless multimedia sensor networks. *Computer networks*, 52(13):2532–2542, 2008.
- [34] Moufida Maimour. SenseVid: A traffic trace based tool for QoE Video transmission assessment dedicated to Wireless Video Sensor Networks. *Simulation Modelling Practice and Theory*, 87:120–137, 2018.
- [35] Cedric Adjih, Emmanuel Baccelli, Eric Fleury, Gaetan Harter, Nathalie Mitton, Thomas Noel, Roger Pissard-Gibollet, Frederic Saint-Marcel, Guillaume Schreiner, Julien Vandaele, et al. Fit iot-lab: A large scale open experimental iot testbed. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pages 459–464. IEEE, 2015.
- [36] Mohammad Rahimi, Rick Baer, Obimdinachi I Iroezi, Juan C Garcia, Jay Warrior, Deborah Estrin, and Mani Srivastava. Cyclops: in situ image sensing and interpretation in wireless sensor networks. In *Proceedings of the 3rd international conference on Embedded networked sensor systems*, pages 192–204, 2005.

- [37] Thiago Teixeira, Eugenio Culurciello, Joon Hyuk Park, Dimitrios Lymberopoulos, Andrew Barton-Sweeney, and Andreas Savvides. Address-event imagers for sensor networks: evaluation and modeling. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 458–466, 2006.
- [38] N Kouadria, N Doghmane, D Messadeg, and S Harize. Low complexity dct for image compression in wireless visual sensor networks. *Electronics Letters*, 49(24):1531–1532, 2013.
- [39] Vitor A Coutinho, Renato J Cintra, Fábio M Bayer, Sunera Kulasekera, and Arjuna Madanayake. A multiplierless pruned dct-like transformation for image and video compression that requires ten additions only. *Journal of Real-Time Image Processing*, 12:247–255, 2016.
- [40] Renato J Cintra, Fábio M Bayer, Vitor A Coutinho, Sunera Kulasekera, Arjuna Madanayake, and André Leite. Energy-efficient 8-point dct approximations: Theory and hardware architectures. *Circuits, Systems, and Signal Processing*, 35:4009–4029, 2016.
- [41] Nasreddine Kouadria, Khaoula Mechouek, Djemil Messadeg, and Nouredine Doghmane. Pruned discrete tchebichef transform for image coding in wireless multimedia sensor networks. *AEU-International Journal of Electronics and Communications*, 74:123–127, 2017.
- [42] Paulo AM Oliveira, Renato J Cintra, Fábio M Bayer, Sunera Kulasekera, and Arjuna Madanayake. A discrete tchebichef transform approximation for image and video coding. *IEEE Signal Processing Letters*, 22(8):1137–1141, 2015.
- [43] Massimo Piccardi. Background subtraction techniques: a review. In *2004 IEEE international conference on systems, man and cybernetics (IEEE Cat. No. 04CH37583)*, volume 4, pages 3099–3104. IEEE, 2004.

- [44] Andrews Sobral and Antoine Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2014.
- [45] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Hélène Laurent, and Christophe Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19(3):33003, 2010.
- [46] Jaya S Kulchandani and Kruti J Dangarwala. Moving object detection: Review of recent research trends. In *2015 International conference on pervasive computing (ICPC)*, pages 1–5. IEEE, 2015.
- [47] Gowrisankar Kalakoti et al. Key-frame detection and video retrieval based on dc coefficient-based cosine orthogonality and multivariate statistical tests. *Traitement du Signal*, 37(5), 2020.
- [48] Thierry Bouwmans, Lucia Maddalena, and Alfredo Petrosino. Scene background initialization: A taxonomy. *Pattern Recognition Letters*, 96:3–11, 2017.
- [49] Fatih Porikli and Oncel Tuzel. Bayesian background modeling for foreground detection. In *Proceedings of the third ACM international workshop on Video surveillance & sensor networks*, pages 55–58, 2005.
- [50] Ashutosh Morde, Xiang Ma, and Sadiye Guler. Learning a background model for change detection. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 15–20. IEEE, 2012.
- [51] Mingliang Chen, Xing Wei, Qingxiong Yang, Qing Li, Gang Wang, and Ming-Hsuan Yang. Spatiotemporal gmm for background subtraction with superpixel hierarchy. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1518–1525, 2017.

- [52] Fida El Baf, Thierry Bouwmans, and Bertrand Vachon. Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 60–65. IEEE, 2009.
- [53] Zhenjie Zhao, Thierry Bouwmans, Xuebo Zhang, and Yongchun Fang. A fuzzy background modeling approach for motion detection in dynamic backgrounds. In *International conference on multimedia and signal processing*, pages 177–185. Springer, 2012.
- [54] Lu Yang, Jing Li, Yuansheng Luo, Yang Zhao, Hong Cheng, and Jun Li. Deep background modeling using fully convolutional network. *IEEE Transactions on Intelligent Transportation Systems*, 19(1):254–262, 2017.
- [55] Jeffin Gracewell and Mala John. Dynamic background modeling using deep learning autoencoder network. *Multimedia Tools and Applications*, 79(7):4639–4659, 2020.
- [56] Mohammadreza Babaei, Duc Tung Dinh, and Gerhard Rigoll. A deep convolutional neural network for video sequence background subtraction. *Pattern Recognition*, 76:635–649, 2018.
- [57] Mohammad Javad Shafiee, Parthipan Siva, Paul Fieguth, and Alexander Wong. Embedded motion detection via neural response mixture background modeling. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 837–844. IEEE, 2016.
- [58] Sandeep Singh Sengar and Susanta Mukhopadhyay. Moving object detection using statistical background subtraction in wavelet compressed domain. *Multimedia Tools and Applications*, 79(9):5919–5940, 2020.

- [59] Ainhoa Mendizabal and Luis Salgado. A region based approach to background modeling in a wavelet multi-resolution framework. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 929–932. IEEE, 2011.
- [60] Majid Sabbagh, Hamed Tabkhi, and Gunar Schirner. Power-efficient real-time solution for adaptive vision algorithms. *IET Computers & Digital Techniques*, 9:16–26, 01 2015.
- [61] Mariangela Genovese and Ettore Napoli. Asic and fpga implementation of the gaussian mixture model algorithm for real-time segmentation of high definition video. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(3):537–547, 2014.
- [62] Kratika Garg, Nirmala Ramakrishnan, Alok Prakash, and Thambipillai Srikanthan. Rapid and robust background modeling technique for low-cost road traffic surveillance systems. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):2204–2215, 2019.
- [63] Jong Hwan Ko, Burhan Ahmad Mudassar, and Saibal Mukhopadhyay. An energy-efficient wireless video sensor node for moving object surveillance. *IEEE Transactions on Multi-Scale Computing Systems*, 1(1):7–18, 2015.
- [64] Nasreddine Kouadria, Khaoula Mechouek, Saliha Harize, and Nouredine Doghmane. Region-of-interest based image compression using the discrete tchebichef transform in wireless visual sensor networks. *Computers & Electrical Engineering*, 73:194–208, Jan. 2019.
- [65] Ahcen Aliouat, Nasreddine Kouadria, Moufida Maimour, and Saliha Harize. Region-of-Interest based Video Coding Strategy for Low Bitrate Surveillance Systems. In *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*, pages 1357–1362, 2022.

- [66] Ahcen Aliouat, Nasreddine Kouadria, Saliha Harize, and Moufida Maimour. Multi-threshold-based frame segmentation for content-aware video coding in wmsn. In *International Conference on Computing Systems and Applications*, pages 337–347. Springer, 2022.
- [67] Jong Hwan Ko, Taesik Na, and Saibal 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*, 8(3):591–602, 2018.
- [68] Jong Hwan Ko. *Resource-aware and robust image processing for intelligent sensor systems*. PhD thesis, Georgia Institute of Technology, 2018.
- [69] Ur Rehman, Muhammad Tariq, and Takuro Sato. A novel energy efficient object detection and image transmission approach for wireless multimedia sensor networks. *IEEE sensors journal*, 16(15):5942–5949, Aug. 2016.
- [70] Huaying Xue, Yuan Zhang, and Yunong Wei. Fast ROI-based HEVC coding for surveillance videos. In *2016 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 299–304. IEEE, 2016.
- [71] Dan Grois and Ofer Hadar. Efficient region-of-interest scalable video coding with adaptive bit-rate control. *Advances in Multimedia*, 2013, 2013.
- [72] Ali Haidous, William Oswald, Hritom Das, and Na Gong. Content-Adaptable ROI-Aware Video Storage for Power-Quality Scalable Mobile Streaming. *IEEE Access*, 10:26830–26848, 2022.
- [73] Odrika Iqbal, Victor Isaac Torres Muro, Sameeksha Katoch, Andreas Spanias, and Suren Jayasuriya. Adaptive Subsampling for ROI-Based Visual Tracking: Algorithms and FPGA Implementation. *IEEE Access*, 10:90507–90522, 2022.

- [74] Zhewei Zhang, Tao Jing, Jingning Han, Yaowu Xu, Xuejing Li, and Meilin Gao. Roi-based video transmission in heterogeneous wireless networks with multi-homed terminals. *IEEE Access*, 5:26328–26339, 2017.
- [75] Wei Jiang and Junjie Yang. Energy-constraint rate distortion optimization for compressive sensing-based image coding. *Signal, Image and Video Processing*, 12(7):1419–1427, 2018.
- [76] Amit Satish Unde and Deepthi P Pattathil. Adaptive compressive video coding for embedded camera sensors: Compressed domain motion and measurements estimation. *IEEE Transactions on Mobile Computing*, 19(10):2250–2263, 2019.
- [77] Jia Guo, Xiangyang Gong, Wendong Wang, Xirong Que, and Jingyu Liu. Sasrt: semantic-aware super-resolution transmission for adaptive video streaming over wireless multimedia sensor networks. *Sensors*, 19(14):3121, 2019.
- [78] Tao Ma, Michael Hempel, Dongming Peng, and Hamid Sharif. A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems. *IEEE Communications Surveys & Tutorials*, 15(3):963–972, 2012.
- [79] Vinay Chande and Nariman Farvardin. Progressive transmission of images over memoryless noisy channels. *IEEE Journal on Selected Areas in Communications*, 18(6):850–860, 2000.
- [80] Vladimir Stankovic, Raouf Hamzaoui, and Dietmar Saupe. Fast algorithm for rate-based optimal error protection of embedded codes. *IEEE Transactions on Communications*, 51(11):1788–1795, 2003.
- [81] Raouf Hamzaoui, Vladimir Stankovic, and Zixiang Xiong. Rate-based versus distortion-based optimal joint source-channel coding. In *Proceedings DCC 2002. Data Compression Conference*, pages 63–72. IEEE, 2002.

- [82] Raouf Hamzaoui, Vladimir Stankovic, and Zixiang Xiong. Fast algorithm for distortion-based error protection of embedded image codes. *IEEE Transactions on image processing*, 14(10):1417–1421, 2005.
- [83] Wei Yu, Zafer Sahinoglu, and Anthony Vetro. Energy efficient jpeg 2000 image transmission over wireless sensor networks. In *IEEE Global Telecommunications Conference, 2004. GLOBECOM'04.*, volume 5, pages 2738–2743. IEEE, 2004.
- [84] Cristina Costa, Fabrizio Granelli, and Aggelos K Katsaggelos. A cross-layer approach for energy efficient transmission of progressively coded images over wireless channels. In *IEEE International Conference on Image Processing 2005*, volume 1, pages I–213. IEEE, 2005.
- [85] Noreen Imran, Boon-Chong Seet, and ACM Fong. Distributed video coding for wireless video sensor networks: a review of the state-of-the-art architectures. *SpringerPlus*, 4:1–30, 2015.
- [86] Ning Ma. Distributed video coding scheme of multimedia data compression algorithm for wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2019:1–9, 2019.
- [87] He Li, Qinglei Qi, Jinjiang Liu, Pan Zhao, and Yang Yang. Mobile wireless multimedia sensor networks image compression task collaboration based on dynamic alliance. *IEEE Access*, 8:86024–86037, 2020.
- [88] MA Matheen and S Sundar. Histogram and entropy oriented image coding for clustered wireless multimedia sensor networks (wmsns). *Multimedia Tools and Applications*, 81(27):38253–38276, 2022.
- [89] Dinh Trieu Duong, Huy Phi Cong, and Xiem Hoang Van. A novel consistent quality driven for jem based distributed video coding. *Algorithms*, 12(7):130, 2019.

- [90] Dongsan Jun. Distributed video coding with adaptive two-step side information generation for smart and interactive media. *Displays*, 59:21–27, 2019.
- [91] Jong Hwan Ko, Taesik Na, and Saibal 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*, 8(3):591–602, 2018.
- [92] Jong Hwan Ko, Burhan Ahmad Mudassar, and Saibal Mukhopadhyay. An energy-efficient wireless video sensor node for moving object surveillance. *IEEE Transactions on Multi-Scale Computing Systems*, 1(1):7–18, 2015.
- [93] Guanlin Wu, Minghai Qin, Tae Meon Bae, Sicheng Li, Yuanwei Fang, and Yen-Kuang Chen. Region of interest quality controllable video coding techniques, 2022.
- [94] Vassilios Tsakanikas and Tasos Dagiuklas. Video surveillance systems-current status and future trends. *Computers & Electrical Engineering*, 70:736–753, 2018.
- [95] Kalpana Goyal and Jyoti Singhai. Review of background subtraction methods using Gaussian mixture model for video surveillance systems. *Artificial Intelligence Review*, 50(2):241–259, 2018.
- [96] Kyungnam Kim, Thanarat H Chalidabhongse, David Harwood, and Larry Davis. Real-time foreground–background segmentation using codebook model. *Real-time imaging*, 11(3):172–185, 2005.
- [97] Olivier Barnich and Marc Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, 20(6):1709–1724, 2010.
- [98] Sandeep Singh Sengar and Susanta Mukhopadhyay. Moving object detection based on frame difference and W4. *Signal, Image and Video Processing*, 11(7):1357–1364, 2017.

- [99] Soharab Hossain Shaikh, Khalid Saeed, and Nabendu Chaki. Moving object detection using background subtraction. In *Moving object detection using background subtraction*, pages 15–23. Springer, 2014.
- [100] Bipin Gaikwad and Abhijit Karmakar. Smart surveillance system for real-time multi-person multi-camera tracking at the edge. *Journal of Real-Time Image Processing*, 18(6):1993–2007, 2021.
- [101] Ruimin Ke, Yifan Zhuang, Ziyuan Pu, and Yinhai Wang. A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on IoT devices. *IEEE Transactions on Intelligent Transportation Systems*, 22(8):4962–4974, 2020.
- [102] Wenhan Yang, Haofeng Huang, Yueyu Hu, Ling-Yu Duan, and Jiaying Liu. Video coding for machine: Compact visual representation compression for intelligent collaborative analytics. *arXiv preprint arXiv:2110.09241*, 2021.
- [103] Wen Gao, Siwei Ma, Lingyu Duan, Yonghong Tian, Peiyin Xing, Yaowei Wang, Shanshe Wang, Huizhu Jia, and Tiejun Huang. Digital retina: A way to make the city brain more efficient by visual coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(11):4147–4161, 2021.
- [104] Lingyu Duan, Jiaying Liu, Wenhan Yang, Tiejun Huang, and Wen Gao. Video coding for machines: A paradigm of collaborative compression and intelligent analytics. *IEEE Transactions on Image Processing*, 29:8680–8695, 2020.
- [105] Yuan Zhang. Video coding for machines. In *ITU Workshop on “The future of media*, 2019.
- [106] Imene Bouderbail, Abdenour Amamra, and Mohamed Akrem Benatia. How would image down-sampling and compression impact object detection in the

- context of self-driving vehicles? In *International Conference on Computing Systems and Applications*, pages 25–37. Springer, 2020.
- [107] Burhan Ahmad Mudassar, Priyabrata Saha, Yun Long, Mohammad Faisal Amir, Evan Gebhardt, Taesik Na, Jong Hwan Ko, Marilyn Wolf, and Saibal Mukhopadhyay. Camel: An adaptive camera with embedded machine learning-based sensor parameter control. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(3):498–508, 2019.
- [108] Zenglu Song, Jin Yao, and Huadong Hao. Design and implementation of video processing controller for pipeline robot based on embedded machine vision. *Neural Computing and Applications*, 34(4):2707–2718, 2022.
- [109] Ahcen Aliouat, Nasreddine Kouadria, Moufida Maimour, Saliha Harize, and Nouredine Doghmane. Region-of-interest based video coding strategy for rate/energy-constrained smart surveillance systems using wmsns. *Ad Hoc Networks*, page 103076, 2022.
- [110] Sandeep Singh Sengar and Susanta Mukhopadhyay. Motion segmentation-based surveillance video compression using adaptive particle swarm optimization. *Neural Computing and Applications*, 32(15):11443–11457, 2020.
- [111] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.
- [112] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.
- [113] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

- [114] Azeddine Beghdadi, Ismail Bezzine, and Muhammad Ali Qureshi. A perceptual quality-driven video surveillance system. In *2020 IEEE 23rd International Multi-topic Conference (INMIC)*, pages 1–6. IEEE, 2020.
- [115] Junhui Zuo, Zhenhong Jia, Jie Yang, and Nikola Kasabov. Moving object detection in video sequence images based on an improved visual background extraction algorithm. *Multimedia Tools and Applications*, 79(39):29663–29684, 2020.
- [116] Wenshuo Gao, Xiaoguang Zhang, Lei Yang, and Huizhong Liu. An improved Sobel edge detection. In *2010 3rd International conference on computer science and information technology*, volume 5, pages 67–71. IEEE, 2010.
- [117] Weibin Rong, Zhanjing Li, Wei Zhang, and Lining Sun. An improved canny edge detection algorithm. In *2014 IEEE international conference on mechatronics and automation*, pages 577–582. IEEE, 2014.
- [118] Juseong Lee, Hoyoung Tang, and Jongsun Park. Energy efficient canny edge detector for advanced mobile vision applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(4):1037–1046, 2016.
- [119] Lei Yang, Xiaoyu Wu, Dewei Zhao, Hui Li, and Jun Zhai. An improved prewitt algorithm for edge detection based on noised image. In *2011 4th International congress on image and signal processing*, volume 3, pages 1197–1200. IEEE, 2011.
- [120] Girish N. Chaple, R. D. Daruwala, and Manoj S. Gofane. Comparisions of robert, prewitt, sobel operator based edge detection methods for real time uses on fpga. In *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, pages 1–4, 2015.
- [121] Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7):577–586, 2002.

- [122] Bianca Silveira, Guilherme Paim, Brunno Abreu, Mateus Grellert, Cláudio Machado Diniz, Eduardo Antonio Ceşar da Costa, and Sergio Bampi. Power-efficient sum of absolute differences hardware architecture using adder compressors for integer motion estimation design. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 64(12):3126–3137, 2017.
- [123] Sandeep Singh Sengar and Susanta Mukhopadhyay. A novel method for moving object detection based on block based frame differencing. In *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, pages 467–472. IEEE, 2016.
- [124] Mehran Yazdi and Thierry Bouwmans. New trends on moving object detection in video images captured by a moving camera: A survey. *Computer Science Review*, 28:157–177, 2018.
- [125] Thomas Huang, GJTG Yang, and Greory Tang. A fast two-dimensional median filtering algorithm. *IEEE transactions on acoustics, speech, and signal processing*, 27(1):13–18, 1979.
- [126] Dongbo Min, Sunghwan Choi, Jiangbo Lu, Bumsub Ham, Kwanghoon Sohn, and Minh N Do. Fast global image smoothing based on weighted least squares. *IEEE Transactions on Image Processing*, 23(12):5638–5653, 2014.
- [127] Wenshuo Gao, Xiaoguang Zhang, Lei Yang, and Huizhong Liu. An improved sobel edge detection. In *2010 3rd International conference on computer science and information technology*, volume 5, pages 67–71. IEEE, 2010.
- [128] Yi [dataset] Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. CDnet 2014: An expanded change detection benchmark dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 387–394, 2014. last accessed on: 18/11/22.

- [129] Antoni B Chan and Nuno Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE transactions on pattern analysis and machine intelligence*, 30(5):909–926, 2008.
- [130] R. Guerrero-Gomez-Olmedo, R. J. Lopez-Sastre, S. Maldonado-Bascon, and A. Fernandez-Caballero. Vehicle tracking by simultaneous detection and view-point estimation. In *IWINAC 2013, Part II, LNCS 7931*, pages 306–316, 2013.
- [131] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 387–394, 2014.
- [132] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly detection in crowded scenes. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1975–1981. IEEE, 2010.
- [133] EC funded CAVIAR project IST 2001 37540 (2004). <https://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>.
- [134] R. Guerrero-Gomez-Olmedo, R. J. Lopez-Sastre, S. Maldonado-Bascon, and A. Fernandez-Caballero. Vehicle tracking by simultaneous detection and view-point estimation. In *IWINAC 2013, Part II, LNCS 7931*, pages 306–316, 2013.
- [135] Lazar Bivolarsky, Mattias Nilsson, Renat Vafin, and Soren Vang Andersen. Compression for frames of a video signal using selected candidate blocks, March 28 2017. US Patent 9,609,342.
- [136] Bodhisattva Dash, Suvendu Rup, Figlu Mohanty, and MNS Swamy. A hybrid block-based motion estimation algorithm using jaya for video coding techniques. *Digital Signal Processing*, 88:160–171, 2019.

- [137] Andreas Doering, Umar Iqbal, and Juergen Gall. Joint flow: Temporal flow fields for multi person tracking. *arXiv preprint arXiv:1805.04596*, 2018.
- [138] Lingchao Kong and Rui Dai. Efficient video encoding for automatic video analysis in distributed wireless surveillance systems. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(3):1–24, 2018.
- [139] Neal R Harvey and Stephen Marshall. Rank-order morphological filters: A new class of filters. In *IEEE Workshop on nonlinear signal and image processing*, pages 975–978. Citeseer, 1995.
- [140] Zoran Zivkovic and Ferdinand Van Der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780, 2006.
- [141] Chris Stauffer and W Eric 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)*, volume 2, pages 246–252. IEEE, 1999.
- [142] Ahmed Elgammal, David Harwood, and Larry Davis. Non-parametric model for background subtraction. In *European conference on computer vision*, pages 751–767. Springer, 2000.
- [143] Zoran Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 28–31. IEEE, 2004.
- [144] Massimo De Gregorio and Maurizio Giordano. Change detection with weightless neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 403–407, 2014.

- [145] M Fatih Savas, Huseyin Demirel, and Bilgehan Erkal. Moving object detection using an adaptive background subtraction method based on block-based structure in dynamic scene. *Optik*, 168:605–618, 2018.
- [146] Arm. Cortex M3 datasheet. <https://iot-lab.github.io/assets/misc/docs/iot-lab-m3/stm32f103re.pdf>, 2018.
- [147] Yiran Shen, Wen Hu, Junbin Liu, Mingrui Yang, Bo Wei, and Chun Tung Chou. Efficient background subtraction for real-time tracking in embedded camera networks. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 295–308, 2012.
- [148] Yiran Shen, Wen Hu, Mingrui Yang, Junbin Liu, Bo Wei, Simon Lucey, and Chun Tung Chou. Real-time and robust compressive background subtraction for embedded camera networks. *IEEE Transactions on Mobile Computing*, 15(2):406–418, 2015.
- [149] Muhammad Umar Karim Khan, Asim Khan, and Chong-Min Kyung. EBSCam: Background subtraction for ubiquitous computing. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(1):35–47, 2016.
- [150] Dong-U Lee, Hyungjin Kim, Mohammad Rahimi, Deborah Estrin, and John D Villasenor. Energy-efficient image compression for resource-constrained platforms. *IEEE Transactions on Image Processing*, 18(9):2100–2113, 2009.
- [151] Sunil Kumar Katiyar and P V Arun. Comparative analysis of common edge detection techniques in context of object extraction. *arXiv preprint arXiv:1405.6132*, 2014.
- [152] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

- [153] Eren Soyak, Sotirios A Tsaftaris, and Aggelos K Katsaggelos. Low-complexity tracking-aware h. 264 video compression for transportation surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(10):1378–1389, 2011.
- [154] Adel A Ahmed. An optimal complexity h. 264/avc encoding for video streaming over next generation of wireless multimedia sensor networks. *Signal, image and video processing*, 10(6):1143–1150, 2016.
- [155] M Collotta, G Pau, VM Salerno, and G Scatá. Wireless sensor networks to improve road monitoring. *Wireless Sensor Networks—Technology and Applications*, pages 323–346, 2012.
- [156] Roberto [dataset] Vezzani and Rita Cucchiara. Video surveillance online repository (visor): an integrated framework. *Multimedia Tools and Applications*, 50(2):359–380, 2010. last accessed on: 18/11/22.
- [157] Liquan Zhao and Shuaiyang Li. Object detection algorithm based on improved YOLOv3. *Electronics*, 9(3):537, 2020.
- [158] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [159] Christoph Loeffler, Adriaan Ligtenberg, and George S Moschytz. Practical fast 1-D DCT algorithms with 11 multiplications. In *International Conference on Acoustics, Speech, and Signal Processing*,, pages 988–991. IEEE, 1989.
- [160] Yukihiro Arai, Takeshi Agui, and Masayuki Nakajima. A fast DCT-SQ scheme for images. *IEICE TRANSACTIONS (1976-1990)*, 71(11):1095–1097, 1988.

- [161] Sourabh Bharti, KK Pattanaik, and Anshul Pandey. Contextual outlier detection for wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, 11(4):1511–1530, 2020.
- [162] Rahul Kumar Verma, KK Pattanaik, Sourabh Bharti, and Divya Saxena. In-network context inference in IoT sensory environment for efficient network resource utilization. *Journal of Network and Computer Applications*, 130:89–103, 2019.