Université Sorbonne Paris Nord

Galilée Ecole doctorale

PhD THESIS

submitted for the degree of

Doctor of University Sorbonne Paris Nord

Speciality : "information science and technology"

defended by

Ismail BEZZINE

2 Jully 2020

Multichannel Processing and Analysis for QoS and Quality driven video surveillance over wireless sensors

Thesis director : **Pr. Azeddine Beghdadi**

Thesis co-director : **Dr. Mounir Kaaniche**

Thesis co-supervisor: **Dr. Saadi Boudjit**

JURY

Titus Zaharia,	Professor, Telecom SudParis	Opponent
Marco Carli,	Professor, Roma Tre University - Italy	Opponent
Ioan Tabus,	Professor, Tampere University - Finland	Opponent
Sule Yildirim Yayilgan,	Associate professor, NTNU - Norway	Examiner
Ahmed Mehaoua,	Professor, University Paris Descartes	Examiner
Azeddine Beghdadi,	Professor, University Sorbonne Paris Nord	Thesis director
Mounir Kaaniche,	Associate professor, University Sorbonne Paris Nord	Thesis co-director
Saadi Boudjit,	Associate professor, University Sorbonne Paris Nord	Thesis co-supervisor

L2TI LABORATOIRE DE TRAITEMENT ET TRANSPORT DE L'INFORMATION UNIVERSITE PARIS 13

Acknowledgments

First of all, I am most grateful to my three supervisors Prof. Azeddine Beghdadi, Dr. Mounir Kaaniche and Dr. Saadi Boudjit for their support, their patience in helping me and their encouragement during my PhD. I feel very lucky to have worked with them.

This thesis would not have been completed without the precious feedback of my thesis committee members Prof. Titus Zaharia, Prof. Marco Carli, Prof. Ioan Tabus, Dr. Sule Yildirim Yayilgan and Prof. Ahmed Mehaoua. I would like to thank them for their constructive comments on my work.

I would also like to express my thanks to my past and current colleagues at the Laboratory of Information Processing and Transmission (L2TI) of Sorbonne Paris Nord university.

Finally, a special thanks goes to Dr. Audace Manirabona, for his help and the constructive collaboration. Again, without the support of my parents, my sisters and my brother, I would never have been able to overcome the difficulties of my life. I greatly acknowledge them for their tremendous sacrifices and continued support during all the years of my study.

Abstract

Intelligent Video surveillance systems are more and more demanding in terms of quality, reliability and flexibility especially those based on Multimedia wireless sensors networks. As much as new systems pay a lot of attention to high-level features such as abnormal event detection, the quality of the video as well as the quality of the network were for a long time neglected. However, due to some natural distortions, inappropriate coding techniques and/or bad network quality, the video quality may be deteriorated making object/event detection very difficult. A main challenging issue of intelligent video surveillance is to improve Quality-of-Experience of the system. In fact, three major factors are involved in the global quality of a video surveillance system which are: the captured video quality, the video coding and the quality-of-service of the network. To this end, it is primordial to assess the video quality in order to decide whether or not an enhancement is needed. On the other hand, the network architecture must comply with the video surveillance requirements and its specificities in order to guarantee a good quality of service.

In this thesis, we focus our interest on video surveillance system's quality. Our objective is to study the quality aspect in new emergent intelligent video surveillance systems. The principal contributions of this thesis are threefold. First, we propose a new stereoscopic image coding techniques based on sparse optimization of non separable vector lifting scheme. In fact, this stereoscopic coding technique can be extended to the context of multiview coding and may offer best coding performance for 3D video surveillance systems. Then, we introduce a new quality-based intelligent video surveillance architecture based on video quality assessment. A video surveillance oriented video quality database is proposed

within this architecture. Finally, a scheduling model based on priority of traffics for multimedia sensors is proposed. The results of this thesis underline the importance of our contributions in the field of intelligent video surveillance. This work does not claim to have completely resolved the problems raised in this thesis. It constitutes a modest first contribution by having explored and analysed the most crucial problems. Nevertheless, it has the merit of having thoroughly analyzed the problems raised by the video surveillance and of having proposed some solutions that remain to be improved in the context of future work. Finally, the provision to the scientific community of a high-resolution video database, unique to our knowledge, presenting different scenarios, is a considerable contribution.

Résumé

Les systèmes de vidéosurveillance intelligente sont de plus en plus exigeants en termes de qualité, de fiabilité et de flexibilité, notamment ceux basés sur les réseaux de capteurs multimédia sans fil. Bien que les nouveaux systèmes accordent beaucoup d'attention à des fonctionnalités de haut niveau telles que la détection d'événements anormaux, la qualité de la vidéo ainsi que la qualité du réseau ont été plus ou moins longtemps négligées. Ces distorsions sont d'origines diverses, dépendant ainsi des conditions d'acquisition, du codage vidéo et/ou la mauvaise qualité du réseau. La qualité de la vidéo est donc fortement conditionnée par ces trois éléments essentiels de la chaine de vidéo surveillance. En effet, cela se repercute inévitablement sur les performances du système de détection et d'identification d'objets, d'événements anormaux et de façon générale l'interprétation de la scène filmée. L'un des principaux défis de la vidéosurveillance intelligente est donc d'améliorer la qualité d'expérience du système. En effet, il est primordial d'évaluer la qualité perceptuelle vidéo afin de de décider d'éventuels pré-traitements ou post-traitements de rehaussement de qualité. D'autre part, l'architecture du réseau doit répondre aux exigences de la vidéo-surveillance et respecter ses spécificités afin de garantir une bonne qualité de service.

Dans cette thèse, nous nous intéressons à la qualité des systèmes de vidéo-surveillance. Notre objectif est d'étudier l'aspect qualité des nouveaux systèmes émergents de vidéosurveillance intelligente. Les principales contributions de cette thèse se situent à trois niveaux. Premièrement, une nouvelle technique de codage d'images stéréoscopiques basée sur l'optimisation des filtres d'un schéma de lifting non séparable est proposée. En fait, cette technique de codage stéréoscopique peut être étendue au contexte du codage multi-vues et peut offrir de meilleures performances de codage pour les systèmes de vidéosurveillance multi-vues. La deuxième contribution porte sur une nouvelle architecture de vidéosurveillance intelligente basée sur l'évaluation de la qualité vidéo. Une base de données de qualité vidéo orientée vers la vidéosurveillance a été ainsi développée pour la première fois, à notre connaissance, pour ce besoin spécifique. Enfin, un modèle de planification basé sur la priorité des trafics pour les capteurs multimédias est proposé. Les résultats de cette thèse soulignent l'importance de nos contributions dans le domaine de la vidéosurveillance intelligente. Ce travail n'a pas la prétention d'avoir résolu complètement les problèmes soulevés dans cette thèse. Il constituent une première contribution modeste en ayant exploré et analysé les problèmes les plus cruciaux. Il a néanmoins le mérite d'avoir analysé à fond les problèmes que soulèvent la vidéosurveillance et d'avoir proposé quelques solutions qui restent à parfaire dans le cadre de travaux futurs. Soulignons enfin, la mise à disposition de la communauté scientifique d'une base de vidéos haute résolution, unique à notre connaissance, présentant différents scénarios constitue un apport considérable.

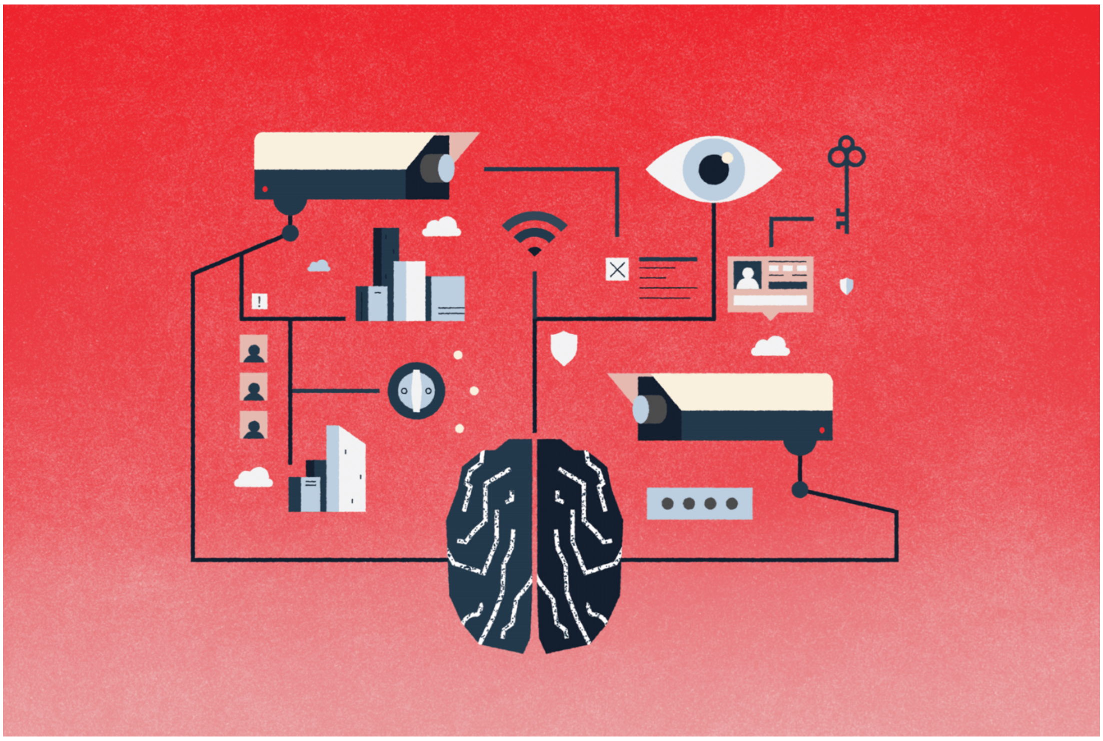
Contents

List of Figures

List of Tables

CHAPTER 1

Introduction



VIDEO surveillance systems are playing an increasingly important role in the remote monitoring of people, property and public and private sites. Their first appearances were in the 1950s. However, video surveillance really developed from the 1970s onwards using closed circuit television (CCTV) systems, mainly in the United Kingdom.

The implementation of video surveillance has intensified during the 1990s. Since the 2001 terrorist attacks in the United States and 2005 in London, the number of surveillance systems installed has increased. Thus, a considerable number of cameras have been widely deployed in public spaces, including transport infrastructures (airports, subway stations,...), parking, banks, shopping malls, roads and industrial sites as a tool for reducing the crime and risk management [1].

Nowadays, video surveillance is one of the oldest and most widespread security solutions. The advent of IP cameras initiated the transition from analog CCTV technology to video over IP networks. This has facilitated the installation of video surveillance networks with a large number of cameras, for example in an airport, hundreds of surveillance cameras can be deployed. These large CCTV infrastructures lead to a huge amount of video streams to be transmitted, viewed and archived. At the same time, the majority of these systems rely heavily on the human operators to monitor scenes and detect suspicious behaviour or events. Unfortunately, many incidents go undetected due to some limitations strongly related to human monitoring capabilities:

- Several screens to be viewed at the same time by the same operator (in practice, each human operator cannot control more than 4 screens at a time [2]).
- Boredom, fatigue and monotony due to continuous hours of monitoring (a 5-10 minute break is recommended every hour for health and safety reasons [2]).
- Ambiguous and unclear information about what we are looking for on the

screen (incidents cannot always be predicted, they can happen unexpectedly). At the same time, abnormal behaviour often triggers suspicion, but it does not always lead to incidents.

- The choice of which camera to look at is made by the operator, making the system vulnerable to abuse (sociological studies state that CCTV officers frequently decide which camera to monitor based on appearance rather than on the behaviour of the people on the screen [3]).
- Other tasks can be managed by the human operator in addition to monitoring (supervising building access, controlling radio communications, etc.).
- The honesty and seriousness of the operators are sometimes questioned.

In large surveillance systems, hundreds of cameras are deployed to ensure absolute control. Authors in [4] have conducted a study in the UK concluding that the screen-for-camera ratio is between 1/4 and 1/30 and the agent-for-screen ratio is around 1/16. Thus, although the fact that theoretically all cameras are monitored, only a small number of screens are monitored in real time. Even these cannot be monitored properly because of the six limitations already mentioned. The rest is saved and viewed later, if necessary, it is then screened and analyzed offline. To overcome these problems, a wave of migration to intelligent surveillance systems is emerging recently, exploiting advanced techniques in video analysis and artificial intelligence. The purpose of using computer vision techniques is to imitate human visual perception and analysis. Although intelligent surveillance systems surpass human capabilities in many cases and human performance is far from optimal, surveillance systems still remain under the supervision of the human agent. However, the superiority of the systems compared to the capacities of the human observer is essentially limited to the aspect of acquisition, parallel processing, the absence of ocular fatigue and optical illusion which could mislead the visual system. On the other hand, it is undeniable that the capacities of scene interpretation by the human being is far superior to that of the machine. An intelligent system based on video would certainly not be able to make the difference, for example, between a scene of a real fight between two individuals and a fake fight, like a scene of hand games. Therefore, exploiting some knowledge of the mechanisms of the human visual system combined with artificial intelligence is a very promising approach. At present, most of the methods for detecting and identifying actions are based on artificial learning [5].

Conventional video surveillance systems are often wired systems that are easy to implement but do not offer either flexibility in camera deployment or the possibility to use removable cameras. In general, these systems are composed of limited number of wired-cameras that are directly linked to the control monitoring center. Fig. 1.1 shows a conventional video surveillance system.

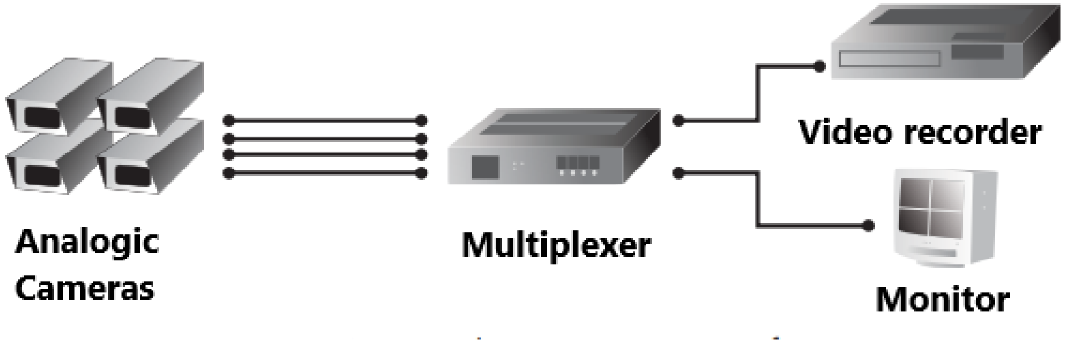


Figure 1.1: *Conventional video surveillance system.*

The recent advancement in micro-electronics technology has opened new perspectives in many fields of applied research including telecommunications, computer science, sensing and imaging. This led to the development of smart sensor devices capable of performing various complex tasks. These smart devices are capable of performing fine grained sensing tasks in a collaborative way through wireless media. This leads to the development of sensor networks dedicated for various applications ranging from large scale habitat monitoring, battle fields observation, and intrusion detection to small critical health monitoring body. Deployed wireless sensor networks used to measure simple physical parameters like temperature, pressure, or humidity, and in general, most of these applications have low bandwidth demands, and are usually delay tolerant. Recently, however, a new application of wireless sensor networks has emerged and it consists of Multimedia Surveillance Wireless Sensor Networks (MSWSNs). This application involves Wireless Multimedia Sensor Networks (MWSNs) and these latters will be composed of interconnected video sensors, each equipped with a low-power wireless transceiver that is capable of processing, sending, and receiving data. Large-scale networks of video and audio sensors could be used to enhance and complement existing surveillance systems and extend the ability of surveillance agencies to monitor areas, public events, private properties and borders. Wireless multimedia sensors could either detect and record potentially relevant activities and make multimedia streams or reports available for future query or send it on real time to a control centre. Wireless sensor networks (WSNs) is consisted of a large number of wireless networked sensors which are a miniature devices that include: a sensing unit for data acquisition, a micro-controller for local data processing, a communication unit to allow the transmission and reception of data to or from other connected motes and a power source which is a small battery in most cases. The collected information is transferred to a central entity called the Sink that is more powerful than the other motes. In fact, this information is analyzed, processed and transmitted via the internet by the Sink. However, the sensor nodes can have different ability, such as different computing power and sensing range (Low or high resolution cameras for Multimedia surveillance). In this case, the network is called heterogeneous wireless sensor network (heterogeneous WSN). Fig. 1.2 shows a Multimedia Surveillance heterogeneous WSN architecture.

One of the most deployed systems over WSNs are security and monitoring systems. Recent years have seen an increase in this kind of systems that control and prevent abnormal events especially in situational awareness applications before irreversible damages occur including national security, deaths and infrastructure destruction. However, despite the tremendous progress already made towards the development of efficient security systems, the existing solutions have limitations especially in complex and cluttered environments such as the environment in a busy soccer stadium or high traffic roads/highways. These difficulties could be alleviated in a Heterogeneous wireless sensor network thanks to the different types of sensors that can be deployed.

Video surveillance systems based on multimedia wireless sensor's network present some major challenges and problems that need to deal with in order to guarantee its efficiency. Therefore, the fast development and widespread deployment of Intelligent video surveillance systems based on Wireless Multimedia sensor networks (WMSNs) especially Wireless Video Sensor Networks (WVSNs) where the sensor is a low cost , low power with good resolution camera and with limited computing resources, has pushed the research community to improve the quality of service by searching for the best technologies to compress the captured information, to manage the channel access/allocation and to assess and enhance the data quality.

The work presented in this thesis is part of a joint international research project involving multidisciplinary teams from Northumbria University (UK), Qatar University and University Sorbonne Paris Nord. This research project aims to develop new solutions incorporating advanced techniques for multisensor signal pre- and post-processing, multimodal data and information fusion, QoS and intelligent sensors connectivity and secure wireless communications. It mainly focuses on some parts where it can be possible to provide innovative solutions to overcome the limitations of the existing video surveillance systems. The project considers some challenging issues related to video surveillance over Wireless Sensor Networks (WSNs) and multi-channel information processing and analysis.

In this thesis, we have mainly focused on the quality aspect for intelligent video surveillance systems based on WMSNs. The global quality scheme is presented in Fig. 1.3. This scheme is composed of three main blocs that treat each of them a particular quality of the system such as the video coding quality, the global video quality and the multimedia wireless sensor network quality.

1.1 Quality of Service in Wireless Multimedia Sensor Networks

Video surveillance over Wireless Sensor Networks requires a high throughput and a high data delivery rate. However, wireless sensor platforms offer limited bit rates (e.g 250 kb/s for IEEE 802.15.4). Therefore, to achieve emerging applications that need higher data rates with low bandwidth and low power operation of the radios,

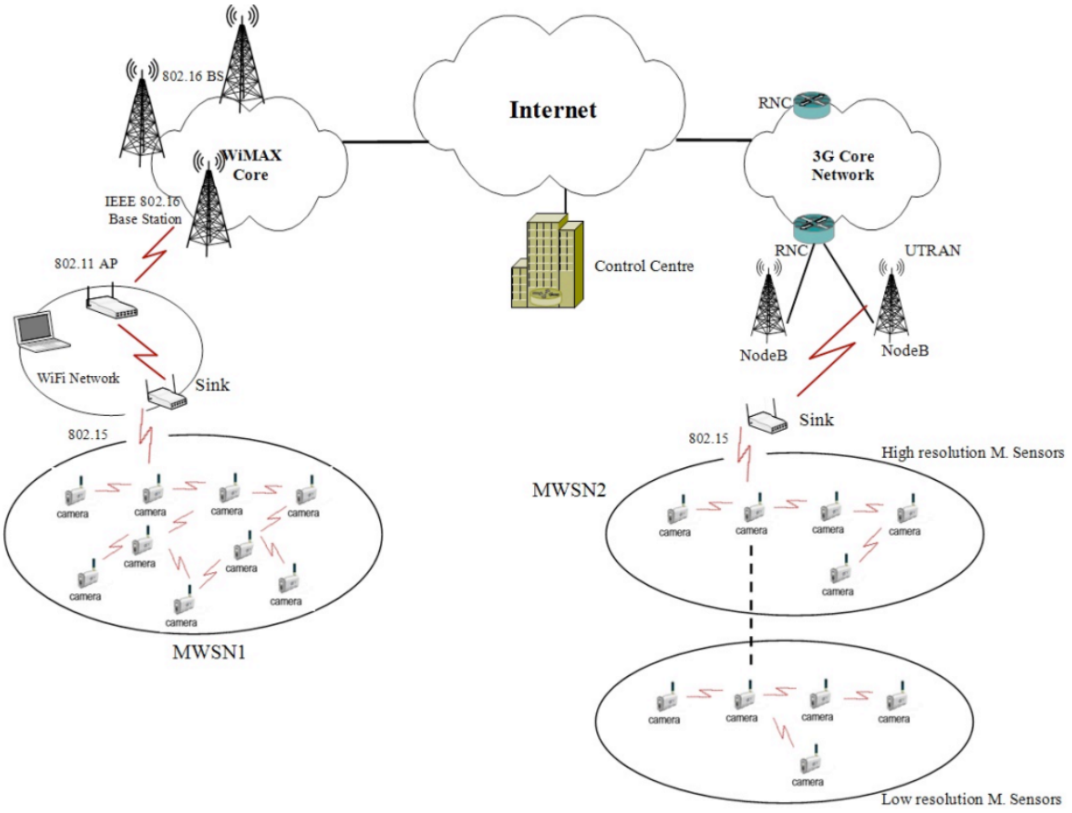


Figure 1.2: *Multimedia Surveillance heterogeneous WSN architecture.*

we need to optimize the channel access by prioritizing the video traffic. In fact, the camera nodes are deployed in different sites and places of different importance, for example in a football stadium the cameras placed in the VIP room and the players entrance are more important than those deployed in stadium borders. Known that the most important factor for video surveillance is the real-time aspect, some information are very delay sensitive, like abnormal event detected. Thus, we need to make sure that the information is transmitted correctly on time. For this reason, we proposed in this thesis a model of scheduling based on prioritization of multimedia data with delay sensitivity.

1.2 Video coding

Multimedia devices especially cameras generate voluminous traffics and even with a network throughput improvement, it is still difficult to transmit the huge amount of data generated by the surveillance sensors. Video coding is one of

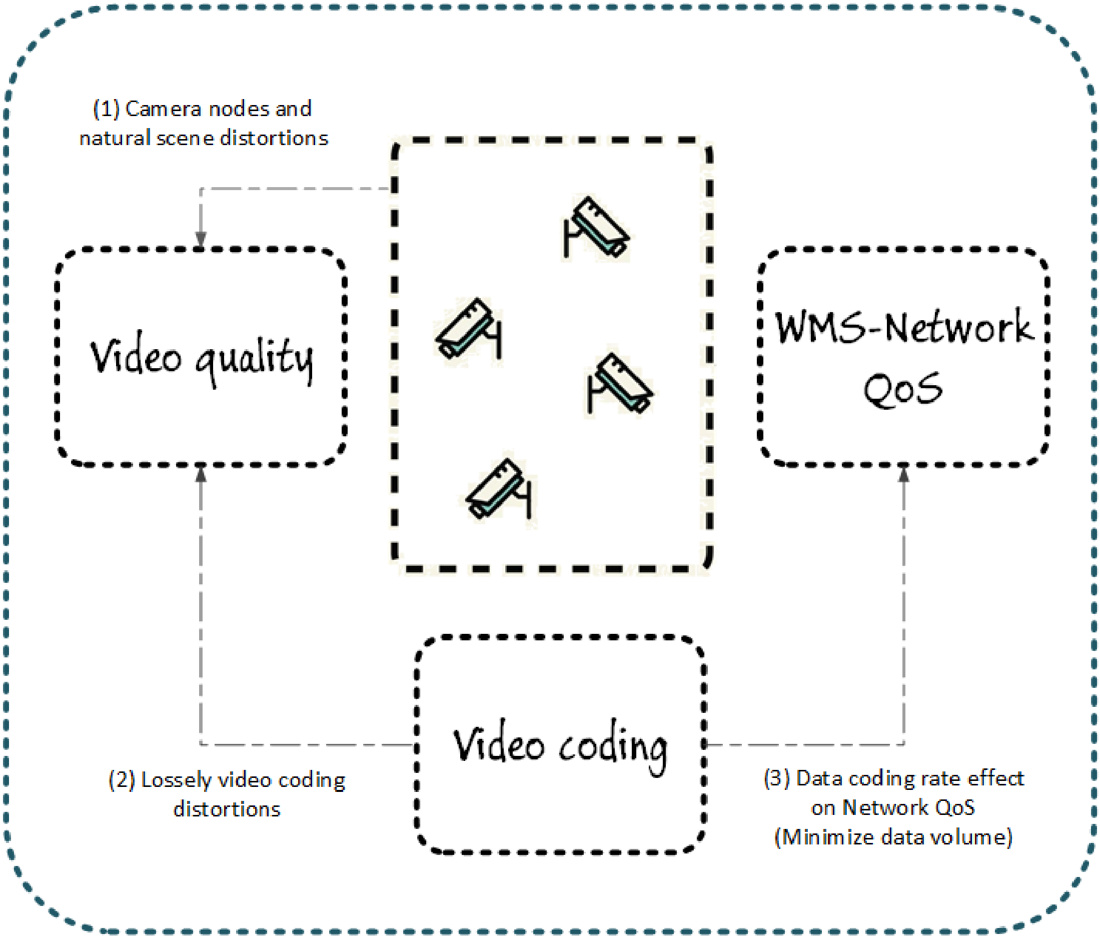

(1) Camera nodes and natural scene distortions
Video quality
WMS-Network QoS
Video coding
(2) Lossely video coding distortions
(3) Data coding rate effect on Network QoS (Minimize data volume)

Figure 1.3: *Video surveillance global quality scheme*

the most important part for an efficient video surveillance system over wireless sensor's network. Many algorithms and codecs have been developed for video coding. However, the increasing interest in stereo images and multiview videos in the last decade has resulted in new Visual surveillance system. In fact, those techniques can provide more details about the captured object and scenes. In particular, a stereoscopic system consists in generating two images by recording the same scene from two slightly different positions. The obtained images, referred to as left and right images, are then merged by the brain to perceive the scene in three dimensions. Despite its advantages in object and scenes detection and understanding, such system generates huge amounts of data which will constitute a problem for its practical use. In fact, this huge amounts of data can affect the efficiency of a visual surveillance system over wireless sensor networks. Therefore, it becomes mandatory to design efficient stereo images /multiview video coding schemes in order to improve their transmission visual quality while reducing their storage capacity.

1.3 Video Quality Assessment

Digital videos are subject to a wide range of distortions arising at different stages of the communication process. These video distortions can have different origins: starting from the video capture to the processing (compression) and transmission phases and ending by the display. In a video surveillance system, the environment can cause some video distortions. In facts, Light levels can change the apparent color and tone of images. In addition to the environment distortions, coding and display artifacts are the most significant types of distortion. Thus, all the above mentioned artifacts and other sources of noise degrade the image quality and result in reducing the accuracy of the different high level tasks like object recognition, abnormal detection, visual tracking and decision.

In order to evaluate the quality of the acquired video as required later for compression, we focus in this thesis on how to measure the visual quality of experience in terms of existing metrics, parameters, and validation procedures. Therefore, the quality assessment will be performed at two levels to help in improving the overall system performance. One of the main novelties in this work is to propose a new video quality assessment database for video surveillance context. To the best of our knowledge, there is no available video database that deals with video distortions for video surveillance applications. In fact, there are a lot of challenges in designing an efficient quality assessment bloc for intelligent visual surveillance video systems.

To sum up, the quality problem in MWSN-based video surveillance systems can be summarized in a closed three inter-dependant blocs loop. Therefore, video coding quality can affect both network and video quality. Thus, it is very important to use a coding scheme that can fulfill the quality requirement of both the video and the network. In the other hand, the video quality assessment is very important as it allows us to evaluate the quality of the video even before transmission, which prevents the network from re-transmitting the same video several times in order to guarantee a good video quality for the abnormal event detection or face detection/recognition process.

1.4 Outline

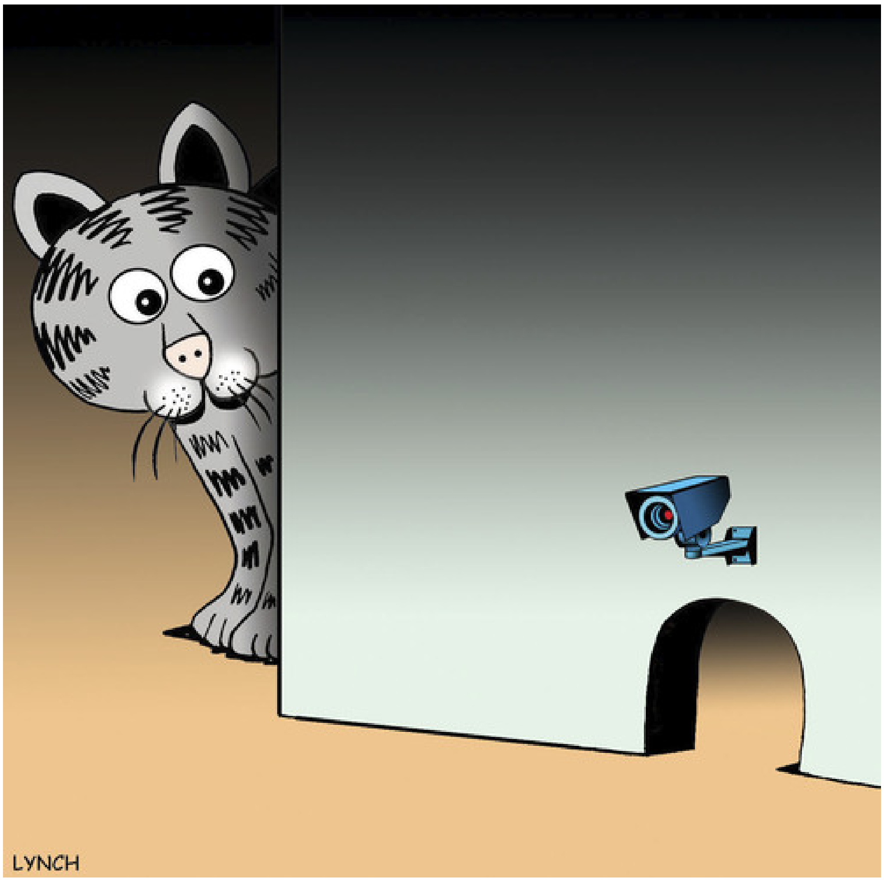
In this chapter, we presented the general context of our work. The rest of this manuscript is organized according to the chronological order of the concerned task in the video surveillance architecture. It is very important to mention that the problem of sensor deployment was not addressed in this thesis. So, the first block in the architecture of the video surveillance system is the capture of the video. Once captured, the video must be encoded to reduce its size. This process is important for the video transmission over the network. Finally, once received, the quality of the video must be evaluated in order to decide whether or not an improvement is necessary. Therefore, the chapters of this manuscript are organized as follows:

- **Chapter 2:** an overview of the intelligent video surveillance systems is given in this chapter.
- **Chapter 3:** This chapter presents a new coding scheme for stereo images that can be extended to multi-view video coding scheme for video surveillance systems.
- **Chapter 4:** It presents a scheduling architecture model based on priority of traffic consisting of a preemptive and non preemptive priority components and a weighted round robin component in order to minimize the delay on the one hand and maximize the throughput on the other hand.
- **Chapter 5:** the importance of the video quality assessment for video surveillance systems and presents a new VQA database for video surveillance context is discussed in this chapter.

- **Conclusion:** This last part is dedicated to concluding remarks and some perspectives and future works. It also highlights the main contributions of this thesis.

CHAPTER 2

Smart video surveillance systems



2.1 Introduction

Video surveillance is a segment of the physical security industry that consists of remotely monitor public or private places with cameras. These cameras are cheap and easy to get, but the manpower required to monitor them is expensive. Therefore, the video traffics from these cameras are generally monitored in moderation or ignored. They are often used as simple archives or to send back an alert once an incident has occurred. Today, the surveillance cameras have become a much more useful tool. Instead of passively record images, they are used to detect events that require a attention at the same time as they occur, and take action in real time. The CCTV for humans is one of the most active research topics in the vision by computer. It has a variety of promising security applications.

2.2 Generalities about video surveillance

2.2.1 Video surveillance market and use

Today, we are witnessing the proliferation of video surveillance in every country in the world. Commercially available video surveillance systems serve several applications: the protection of sensitive sites (government buildings, nuclear power plants, river dams), public places (museums, airports, train stations, banks, shopping malls, stadiums), home security (theft detection, fire detection), surveillance of the elderly (activity analysis, fall detection), road safety (flow estimation, air traffic control, accident detection), detection of abnormal events (detection of cheating in schools, detection of crime in cities), industrial safety (surveillance of workers in factories, access control), etc [6].

According to the IHS Technology study, 1.2 billion video surveillance cameras has been installed for professional purposes will are active and operational worldwide in 2018. The top 5 cities in the world with more surveillance cameras are respectively: Chongqing (China), Beijing (China), London (UK), Chicago (USA), Houston (USA), New York City (USA). National security, having become a very sensitive issue, is the most important factor of the increases of the video surveillance systems market.

2.2.2 The most common video surveillance systems

The need for video surveillance has led to the launch of various major research projects. Several of them have proven their effectiveness and have become widespread global solutions, such as the following:

- The Video Surveillance and Monitoring (VSAM) system that automatically analyzes the activities of objects in ordinary battlefield or civilian scenes [7].
- the Pfinder system that accurately tracks a moving person in complex scenes
- IBM's Smart Surveillance Solution (S3), which enables detection, tracking and classification of objects according to facial colour [8].
- the W4 real-time surveillance system, which uses a combination of shape analysis and tracking, and builds models of people's appearances to detect, track groups of people in occlusion and monitor their behaviour [9].
- the HID (Human Identification at a Distance) system which classifies and identifies human beings at great distances.
- the European ADVISOR project which is a basic project on surveillance in metro stations.
- the VIEWS road surveillance system which plays a very important role in traffic control [10].
- Smart Catch used in San Francisco International Airport can detect anomalies or suspicious behaviour.
- IVA software (Intelligent Video Analysis) which ensures the surveillance of Athens International Airport.

2.3 Conventional video surveillance architecture

In this section, the different hardware and software components of video surveillance systems are presented in summary form. As shown in Fig. 2.2, common video surveillance systems generally consist of the following stages: acquisition, compression, transmission, decompression and processing.

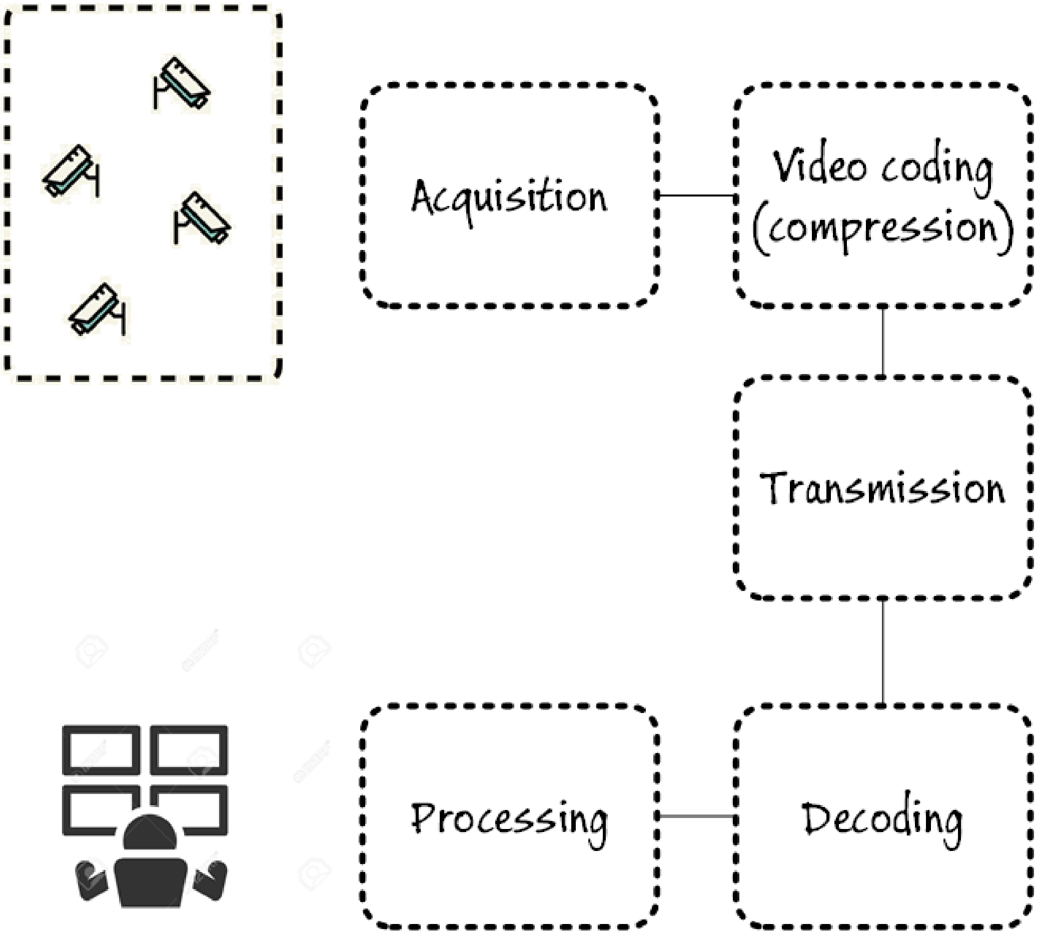


Figure 2.1: *Conventional video surveillance architecture*

- **Acquisition** : The scene to be monitored is recorded by a surveillance camera. There are a variety of camera models to meet different surveillance needs. They are analog or digital and can be motorized or not.

- **Coding (Compression)** : The digitized video sequence represents a large amount of data to be transmitted and processed. This requires a fairly large bandwidth and storage space. Since this is not always available, the video must be compressed in order to reduce the amount of data by removing redundancies between images as well as details imperceptible to the human eye.

- **Transmission** :The video sequence captured by the surveillance cameras must be transmitted to the control unit. Several support of transmission are provided.
- **Processing** : Upon arrival at the control unit, the video streams may undergo different processing, depending on the purpose of the surveillance system application, such as recording, viewing, analysis and searching of the recorded sequences. Some systems simply archive the video sequences for a limited period of time. The recordings are only viewed when necessary. Others provide direct supervision of hundreds of cameras by human operators. Intelligent video surveillance systems automatically analyze the video sequence in real time the transmitted scenes and alert the operator in case of suspicion.

2.4 Video surveillance systems classification

A wide variety of monitoring systems are offered to date. These systems can be classified according to different criteria.

2.4.1 Classification according to automation level

After exploiting advanced video analysis and artificial intelligence techniques, video surveillance systems can be classified into three types: manual, semi-autonomous and fully autonomous [1].

- **Manual video surveillance systems** : They require a human operator monitoring the screens directly. These systems are still widely used.

- **Semi-autonomous video surveillance systems** : They combine video processing and human intervention. Take the example of a system where only unexpected movements are recorded and sent for analysis by a human expert.

- **Autonomous video surveillance systems** : They are also called intelligent video surveillance systems. They can provide reliable real-time monitoring by intelligently analyzing video data without human intervention. Intelligent systems must meet three important characteristics: (1) operation without human control; (2) prediction of events, behaviours, movements; and (3) monitoring, control and alerting of unanticipated activities.

2.4.2 Classification according to the network architecture

Video surveillance systems can be deployed according to two main types of architecture, either centralized or distributed [1].

- **Centralized architecture** : In a centralized architecture, all processing is performed in the same control station. The encoding, recording, viewing and analysis of video streams require a great deal of computing power. In addition, the transmission of all video streams to a centralized point consumes a lot of bandwidth.

- **Distributed architecture** : In a distributed architecture, processing is distributed to the different nodes of the video surveillance system. Thus, the calculations necessary for analysis can be made on intelligent cameras equipped with processors, or in the encoders. This architecture reduces the necessary bandwidth and facilitates the extension of the camera network since the addition of cameras does not affect the computing power of the end station.

2.4.3 Classification according to field of application

Video surveillance applications can be divided into five categories according to their objectives [11].

- **Protection and confidentiality** : Video surveillance is massively deployed for the protection of people and places. It is widely used by governments for internal security [12], and the security of public sites (museum, airport, train station, bank, ...) [13], [14]. It is also used for home surveillance [15], [16], surveillance of the elderly [17], [18], ...

- **Object analysis** : Some video surveillance systems are useful for discovering the trajectories of people through tracking [19], [20]. Others help to monitor complex environments such as supervising workers' activities in factories [21], estimating the length of the queue [22] or even in autonomous navigation.

- **Object recognition** : It encompasses all video surveillance applications where the identity of moving objects (pedestrians or vehicles for example) is revealed by the detection of characteristic elements such as: face recognition [23], [24], license plate recognition, classification of moving vehicles [25]. Also, systems where the behaviour of moving objects can be analysed, recognised and interpreted [26], [27].

- **Traffic monitoring** : Automatic traffic monitoring plays an essential role in road traffic control. Advanced systems facilitate the management, safety and analysis of traffic in road networks such as: estimation of vehicle flow rates, vehicle speed control, calculation of traffic density on the motorway [28], [29], air traffic control [30] and maritime traffic control [31], [32].

- **Abnormal event detection** : The purpose of these systems is to monitor environments, detect abnormal events and alert in some cases, such as:

fire detection [33], accident detection [34], crime detection [35], cheating detection [36],...

2.4.4 Classification based on the application's purpose

Intelligent video surveillance systems can provide results ranging from low-level such as object detection to very advanced levels such as object behavior analysis. These results are highly dependent on the system requirement. Depending on the desired result, video surveillance systems can have different levels of processing. Hierarchically, they start from the pixel level, in through objects to reach the behavioural scale. Authors in [37], have proposed a four levels classification based on the complexity application of the system. In facts, there are four main tasks that the system can perform depending on its application level: object detection, object tracking, object classification and identification, activity analysis and object behavior analysis.Thus, video surveillance systems can be grouped into four categories according to the tasks that it can performs.

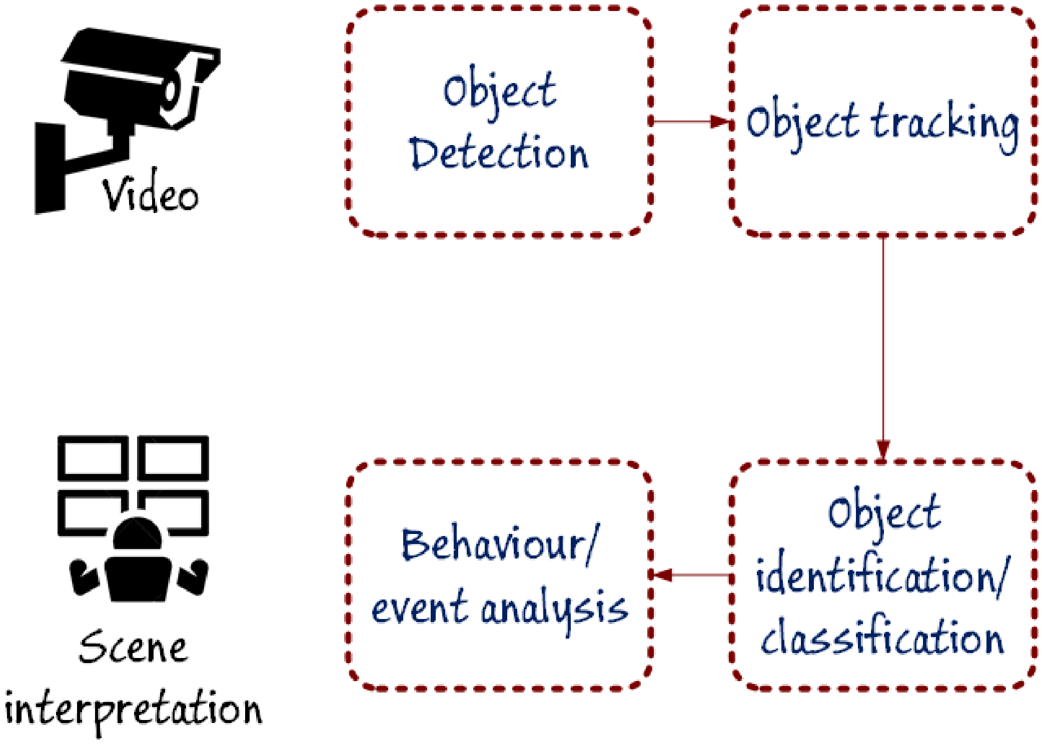

Video
Object Detection
Object tracking
Object identification/ classification
Behaviour/ event analysis
Scene interpretation

Figure 2.2: *Four main tasks achieved by video surveillance system*

- **First level** : This category includes low-level applications for video surveillance systems. The object detection functionality is sufficient for these simple applications. In most systems, the used cameras are assumed to be static. Detection and/or counting applications are mainly based on the object detection feature without the need to reach higher level functionalities. These applications are used to count the number of people entering and/or leaving a building [38] [39], alert when an activity is detected in a scene, estimate the length of queues in shops, monitor bus terminals or train stations [40]. Detection and vehicle counting is necessary to calculate traffic congestion, estimating vehicle throughput and keeping track of vehicles that follow a particular route [41].

- **Second level** : The object tracking task, preceded by its detection, is used by the intermediate level (second level) of video surveillance applications. It aims to detect the trajectory of a moving object in the scene. Applications belonging to this category are used for monitoring traffic congestion [42], detection of abnormal events [43], detection of disputes, detection of abandoned or stolen objects [Antonio 2013], monitoring of elderly people [44], detection of presence in a restricted area, parking of cars and sudden stop of moving objects [43]. [45].

- **Third level** : These applications refer to classification and/or identification systems. The tracking information associated with various extracted attributes (colour, size, etc.) is the key to automatically analysing the objects (type, identity, etc.). Objects detected by a video surveillance system are generally classified into different categories: human, vehicle, animal, etc [46]. The identification of the detected object is directly related to its class. In facts, It is facial if the object is a human being, or based on the analysis of the plate registration if the object is a vehicle.

- **Fourth level** : Analyzing and understanding behaviour is considered as the highest level task used in video surveillance applications. It is an essential step in which information from lower-level functionalities is combined and interpreted through high-level semantic description. The behavioural analysis function is useful in exceptional event detection applications, such as the detection of suspects or missing persons [47], detection of cheating [48], surveillance of the elderly [49], etc. It is also reliable for crowd analysis [Simone 2014] and traffic analysis [50].

2.5 Intelligent video surveillance systems

With the massive deployment of video surveillance cameras, the video stream to be archived becomes colossal, which exceeds the capacities of the surveillance agents. To process all this information, intelligent software has been developed for automatic scene analysis. Hence the emergence of the concept of intelligent video surveillance, which offers systems that allow objects to be detected and tracked and suspicious events to be reported. Thus, It offers solutions for processing recorded video sequence in order to retain only significant and important information. Detection, tracking, classification and identification, as well as behavioral analysis of moving objects are now well-established techniques in intelligent video surveillance. The processing flow is almost always unidirectional starting from the object detection to finish with the behaviour and/or activity analysis. Fig. 2.3 shows a block-diagram of intelligent video surveillance systems.

2.5.1 Moving object detection

Object detection is usually the basis of any intelligent video surveillance system. It detects activities in the monitored scene, such as the movement, appearance or disappearance of an object. Object detection is aligned with motion detection since the moving parts of the scene are the regions of interest (foreground) and the static parts are not (background) [12]. Many motion detection techniques are based on

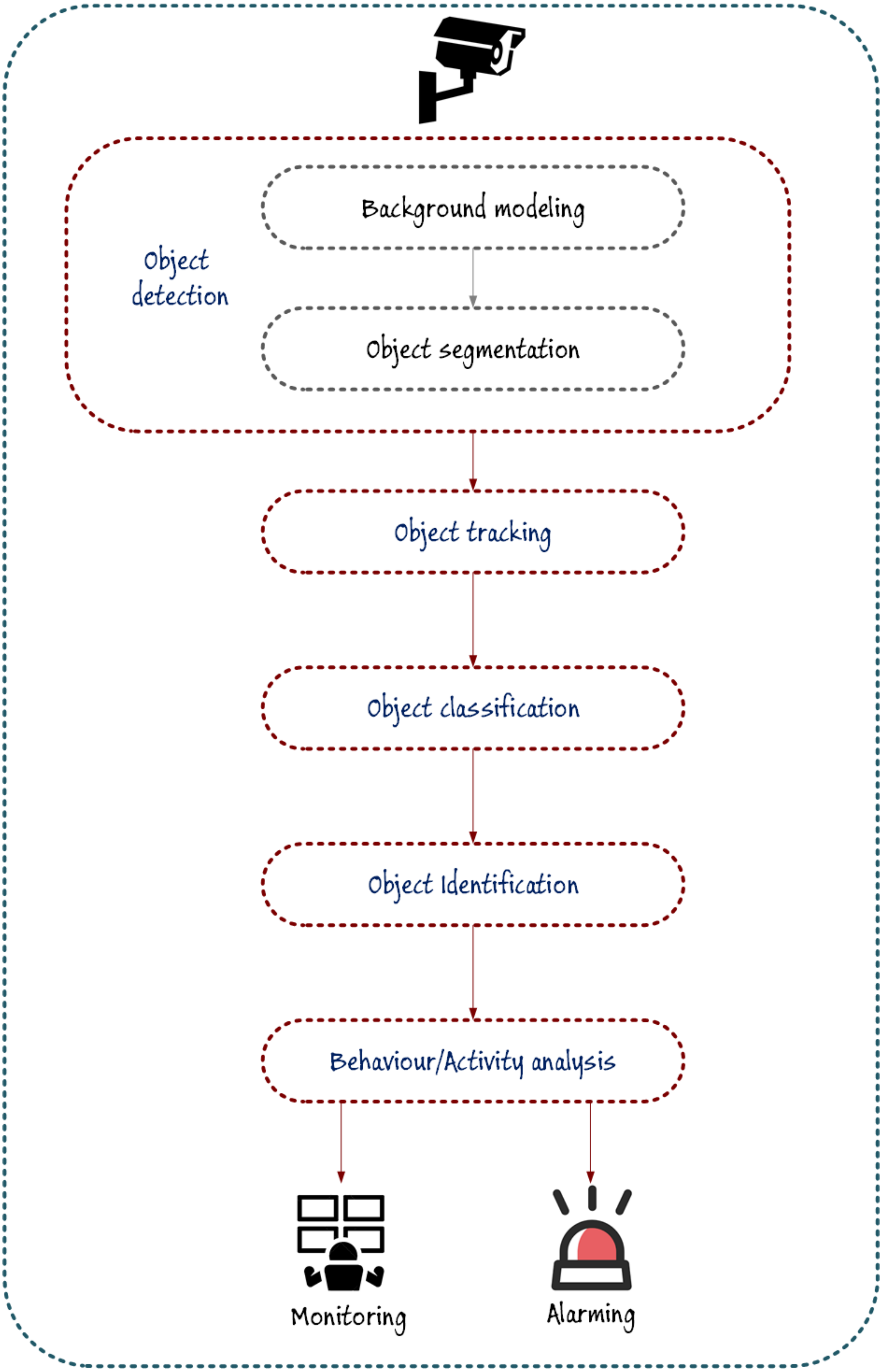


Figure 2.3: *Intelligent video surveillance systems block-diagram*

change detection. However, detecting changes in a scene may not necessarily target the movement of objects, but it can highlight a modulation of the image. In order to segment moving objects, we must be able to make the difference between pixels that correspond to consistent motion and those caused by environmental changes. Complex environments can present a major problem due to many variations (lighting changes, unnecessary movement, cluttered backgrounds).Several motion segmentation techniques are commonly used in the literature such as:

2.5.1.1 Temporal difference

A first class of motion detection methods is based on temporal difference between images. They do not require background models. In facts, they extract the moving regions by analyzing the temporal variation of the light intensity of the pixels. They are very fast and adaptable to dynamic environments.

2.5.1.2 Background subtraction

Background subtraction consists of two main steps: (1) Modeling the background; (2) Motion segmentation. Background modeling is the representation of the scene without the moving objects and must be updated regularly. The motion segmentation aims at detecting regions corresponding to moving objects (people, vehicles, ...). The video frames are compared to the background model and the differences are marked as moving objects [51].

2.5.2 Moving object tracking

Once moving objects are detected, their movements are tracked throughout the video sequence. Tracking is the estimation of the trajectory of an object in the image plane as it moves through the scene. This task requires locating each object in every video frame. Tracking can be done in 2D, from a single camera, or 3D, by combining two views with a known geometric relationship. Many tracking techniques predict the position of the object in a frame based on its movements observed in previous images. Each detected object must be associated with its correspondent in the next frame to update its trajectory, otherwise a new trajectory is created. Tracking these objects can be difficult due to the complexity of their shapes, their non-rigid nature, their movements, partial or complete occlusions, changes in scene lighting, etc. These can be simplified by simple assumptions such as smooth movements, and prior knowledge of the number, size, shape and appearance of the objects. Tracking allows extracting other characteristics: trajectory, speed, direction of movement, position at a specific time. There are different classifications available for object tracking methods, such as [52], [53], [54]. The most popular classifications are [54] which classifies object tracking algorithms into four categories: region-based algorithms, active contour-based algorithms, feature-based algorithms, and model-based algorithms, and [52] where the authors classify algorithms into three categories: point tracking, kernel tracking, and silhouette tracking.The most exhaustive classification is proposed in [55] where authors classify the existing tracking methods into four categories: tracking based on the matching, filter-based tracking, filter-based tracking, filter-based tracking, filter-based tracking the class and follow-up based on the merger.

2.5.3 Moving object classification

Classification is an object recognition task. In order to track them and analyze their behavior, it is essential to classify them correctly. Detected objects can generally be classified into vehicles, animals, humans and other moving objects [46]. In general, the system recognizes the nature of a detected entity from the attributes of its shape and/or the properties of its movement [12]. Classification approaches are based on motion, shape, color and texture.

2.5.3.1 Shape-based approaches

Shape-based classification is strictly concerned with the geometry of the object. Depending on the geometry of the extracted regions, such as enclosing boxes and external contours, objects can be classified. The authors of [56] explore the study of various characteristics of the shapes with precision. Tsai et al [57] present a method to track humans in crowded scenes, using the models of human forms, in addition to camera models. According to [58], the classification based on on the form has reasonable accuracy. Its calculation time is considered low compared to to other methods of classification.

2.5.3.2 Motion-based approaches

The motion-based approach provides a robust method for classification [59]. It does not require predefined shape models, but it has difficulties in identifying a non-moving object [60]. Although motion-based classification has moderate accuracy, it does not require much calculation resources. Non-rigid moves of articulated objects have a very interesting periodic property for their classification. Optical flow is also very useful: the residual flow can be used to analyze the rigidity and periodicity of moving entities. Rigid objects are expected to have less residual flux than non-rigid ones [61]. Johnsen et al [62] propose a tracking and classification system that has shown good results on multiple objects under various light and occlusion conditions.

2.5.3.3 Color-based approaches

Unlike many other image characteristics (e.g., shape), color is relatively constant during changes in viewpoint and is easy to acquire. The representation of color characteristics is the most effective way to reveal the similarity in color images. In content-based image search systems [63], the simplest and most effective searches are those based on color. The authors in [64] propose a moving object tracking system based on segmentation of color images and color histograms. According to [60], the accuracy and computation time are high for classification based on on the color.

2.5.3.4 Texture-based approaches

Assigning an image to a known texture class is an important objective of texture-based classification. With the existence of several classifiers, the main task is to extract the relevant features from the textured image. These approaches consist of two phases: the learning phase and the recognition phase. Texture-based methods such as HOG (Histogram of Oriented Gradient) histograms use contour-based dimensional features [65]. In accordance with [60], these methods give better accuracy but with additional computational time.

After determining the class to which an object belongs, its identity must be revealed. In surveillance systems for access control or the search for suspects [21], in addition to classification, the identity of the object must also be revealed, for example, by facial recognition of the individual or by reading the car license plate. A lot of research has been conducted in recent years in these two specialized applications. Facial recognition is one of the main tools used for biometric identification of people on video [66] because it allows more accurate identification. However, facial recognition in an uncontrolled environment remains a problem that has not yet been satisfactorily resolved. Reading license plates in video surveillance systems is a difficult application [67]. In facts, It requires a high-resolution image. Image analysis is confronted with many environmental interferences. To maximize efficiency, license plate recognition is most often performed by specialized systems with well positioned cameras and adequate lighting quality.

2.5.4 Behavioural analysis of objects in movement

Behaviour analysis is the highest level task used by intelligent video surveillance systems. The information collected by the previous steps is interpreted through semantic description to describe the behaviours and interactions of objects in the scene with natural language. Semantic analysis is often highly dependent on the context of the application. The most commonly used techniques to model the detected behaviours are: Hidden Markov models, neural networks, Bayesian networks [54], etc. First, visual information of moving objects in the scene is extracted and described with an appropriate method, then these information are studied to recognize and understand behaviour. Many characteristics have been proposed to describe human activities based on three main algorithms [68].

2.5.4.1 Algorithms based on 3D models

The most common technique to acquire the 3D information of a movement is to retrieve the pose of the person or object at each moment using a 3D model. The model is constructed by trying to minimize a residual measurement between the projected model and the contours of the object. This usually requires a strong foreground/background segmentation. As an example, Campbell and Bobick [69] have calculated 3D information of the positions of the human body parts. Their system exploits redundancies that exist for particular actions and performs recognition using only the information that varies between actions. This method only examines the relevant parts of the body.

2.5.4.2 Algorithms based on appearance models

Unlike 3D algorithms, other works try to use only the 2D appearances of the action. An action is described by a sequence of 2D instances/positions of the object. Many methods require a standardized image of the object (usually without background). For example, Cui et al [70], Darrell and Pentland [71], and also Wilson and Bobick [72] present results using actions (mainly hand gestures), where grayscale (backgroundless) images are used. Although hand appearances remain quite similar in many people, with the obvious exception of skin color, actions that include the appearance of the whole body are not visually consistent for different people due to natural variations and different clothing appearances.

2.5.4.3 Algorithms based on motion models

These approaches attempt to characterize movement without reference to static body postures. The authors of [73] use repetitive motion as a strong warning signal to recognize cyclic walking movements. They track and recognize people walking in outdoor scenes by collecting a vector that characterizes the whole body. This vector carries low-level movement characteristics and periodicity measurements. Other work, such as [74], focuses on movements associated with facial expressions using movement properties based on predefined regions. The goal of this research is to recognize human facial expressions as a dynamic system, where the movement of regions of interest is relevant. These approaches characterize expressions using the properties of underlying movements rather than representing the action as a sequence of underlying movements poses.

After characterizing the behavior, its patterns are analyzed to be recognized. At present, the recognized behaviours are mainly: head and limb movements and gestures. There are two types of behavior recognition algorithms, as follows:

- **Template Matching Method** : The basic idea is to extract characteristics from the video sequences and then compare them with pre-recorded behaviour patterns. This method has a low computational cost, but is sensitive to noise.

- **State-space method** : Each static gesture is defined as a state, then all these states are combined with a probability. Each behavior is considered as a set of states. The classification of the behavior depends on the maximum value of the joint probability. This method requires a complex iterative calculation.

CHAPTER 3

Visual data coding

Abstract

Video-based surveillance and in particular multi-sensor-based systems have been undergoing massive development and deployment in various applications and needs over the last few years. The recognized successes make it the most reliable solution for the security of public places and people. Thanks to its tremendous advantages, multiview based video surveillance systems are more and more employed to cope with classical systems limits such as occlusion problem especially for crowded scenes. The great interest in these systems has resulted in huge amount of data which needs to be compressed for storage and transmission purposes. In this thesis, we focus on the stereoscopic coding issue as it is a particular case of the general multiview video surveillance system. It is considered as a first step for understanding and solving the challenging issues that may arise when coding multiple views. Therefore, a new stereoscopic view coding scheme that can be generalized to the multiview case is introduced and discussed in the context of video-surveillance. In this context, vector lifting scheme has been found to be an efficient approach for stereo image coding. For instance, the coding performance depends on the design of the involved lifting operators referred to as prediction and update filters. For this reason, while a non separable vector lifting structure is retained, we investigate different techniques for optimizing sparse criteria to design the filters used with both views. More precisely, an independent full optimization algorithm as well as a joint algorithm will be developed and studied. Simulations performed on different stereo images demonstrate the effectiveness of the proposed sparse optimization algorithms in terms of quality of reconstruction and bitrate saving. [75].

[75] I Bezzine et al. “Sparse optimization of non separable vector lifting scheme for stereo image coding”. In: *Journal of Visual Communication and Image Representation* 57 (2018), pp. 283–293

3.1 Introduction

As it was previously mentioned, this thesis is a part of an international project that aims to design and implement new innovative smart video surveillance system (VSS) providing a better scene understanding thanks to some performing abnormal events detection techniques. To this end, we have given more interest to a new video surveillance (VS) framework integrating the classical approaches and new solutions. We have then developed an approach combining different aspects of conventional video surveillance systems with technological advances, and in particular the strategic deployment of wireless sensors and the notion of intelligent video surveillance through the integration of the video quality aspect in this rather particular context allowing to perform optional pre-processing conditioned by the level of the required quality that is decisive for higher level tasks. As some existing video surveillance systems, we consider the multi-view architecture in the proposed solutions. This study looked at the classical case of stereo-vision as a starting point and demonstration of the contribution of the multisensor approach in the context of VS. Our first contribution focused on the coding part which represents one of the essential links in any visual information transmission system. We have taken great care in how to code stereo content efficiently. Unlike conventional 2D video surveillance, multi-view vision methods provide better detection and recognition of the content of the observed scenes, such as human body and any other object shape which facilitates the abnormal event detection process and scene interpretation in VSS.

3.1.1 3D-based video surveillance in WMSN

In the scientific community working in the field of coding of visual content, the concept of 3D coding is used more or less erroneously. It is in fact multi-view information combined with depth maps. The exploitation of these two pieces of information makes it possible to generate visual contents offering a perceptual sensation of volume to the observer. The 3D aspect here has nothing to do with 3D in the field of 3D medical imaging or computer graphics for example. The concepts of muliview coding or 3D coding will therefore be used interchangeably in the following, as is the case in several works published in the field of advanced video coding. The fast pace of technological progress in 3D wireless node camera has lead to a new emergent 3D-based smart video surveillance systems which offer an automatic detection and identification of suspicious objects or critical situations. These systems have appeared thanks to the considerable progress in video surveillance related research areas such as face recognition [76], tracking people person identification and background subtraction [77, 78], posture and gesture recognition [79, 80, 81] and human activity recognition [82]. The most challenging problem for conventional 2D video surveillance systems is the presence of shadows, variations in illumination, cluttered background, and occlusions. Where most of these problem result on a distorted video which can be enhanced using existing video quality enhancement techniques, the occlusion problem leads to a loss of crucial information that can not be easily compensated at the reception. Therefore, 3D video surveillance systems handle the occurrence of occlusions by exploiting the different viewpoints. Moreover, 3D system enables progress in solving complex intelligent video surveillance problems [83] like unusual events detection [84], human–human interaction recognition and prediction [85], abnormal event detection [86, 87, 88] and aggressive behaviour/anger detection [89].

3.1.2 Multiview video coding in WMSN

The multi-view video sensor nodes in video surveillance applications using Multimedia wireless sensors network represent an array of video camera nodes that are arranged in such a way that multiple views of the single scene are captured using multiple camera nodes making the scene clearer. However, the resulting videos from the different views (nodes) contain an enormous amount of redundant data. Hence, complexity arises due to processing, storage and transmission of these huge volumes of data requiring a low complex compression and encoding technique in the encoder side. Basing on the fact that nodes in WMSN have some limitations like memory and computational speed, power consumption and low data-rate, video coding is still one of the most challenging task for such systems.

To this respect, many research efforts have been devoted to the development of efficient multiview video coding schemes. More precisely, a first category of methods, referred to as Distributed Video Coding (DVC), can be found in the literature [90]. The objective of such methods consists in shifting the complexity of exploiting intra and inter frame redundancies from the encoder to the decoder side. The second category of methods, which have attracted more attention, focused on the conventional sources coding schemes by exploiting such redundancy in the encoder side to to achieve better bitrate saving and quality of reconstruction. To achieve this goal, the developed methods aim to improve the motion and/or disparity estimation step as well as the decomposition employed to encode the different signals in the transformed domain (discrete cosine transform and discrete wavelet transform domains).

It should be noted here that our proposed coding method belongs to the second category of the aforementioned methods. Our objective is to find an efficient *joint decomposition* that allows us to provide compact and sparse representation of the different generated views. As mentioned, earlier, and for the sake of simplicity, this approach will be developed and studied by considering only two views (i.e. stereoscopic systems.

3.2 Stereoscopic image coding

A stereoscopic imaging system consists in generating two views, referred to as left and right images, by recording the same scene from two slightly different view positions. The obtained images, referred to as left and right images, are then merged by the brain to perceive the scene in three dimensions. For this reason, stereovision has been widely used in various application fields such as 3DTV, digital 3D cinema, computer vision, remote sensing, visual surveillance and medicine [91, 92]. Thus, the increasing interest in stereo systems especially for Wireless Multimedia Sensors Network for video surveillance purpose has resulted in huge amount of data which will constitute a problem for its practical use. Therefore, it becomes mandatory to design efficient stereo image coding schemes with high visual information transmission quality and low storage capacity.

3.2.1 State-of-the-art methods

A basic approach for stereo images coding may consist in encoding independently the left and right views by employing existing still image encoders. However, such approach may not appear so efficient since it does not exploit the main characteristics of these images. Indeed, as the stereo images correspond to the same 3D scene, they present similar contents and exhibit a high correlation. Therefore, efficient stereo image coding schemes could be designed by exploiting the inter-view redundancies. To this respect, the conventional scheme can be described as follows. First, one image, for example the left one, is selected as a reference image, and the other one (i.e the right image) is considered as a target image. Then, the target image is predicted from the reference one thanks to the disparity estimation/compensation (DE/DC) process. The difference between the original target image and the predicted one leads to the generation of the residual image. Finally, the reference and residual images as well as the disparity information are encoded. It is important to note here that this idea is behind most of the existing stereo image coding methods.

However, they differ in some aspects and could be roughly classified into two categories. The first category of methods aims to improve the DE/DC process as well as the coding of the disparity (or depth) maps [93, 94, 95, 96, 97, 98]. For instance, while the standard block-matching (BM) technique is often used to perform the DE/DC step, modified BM [93] and learning dictionary-based techniques have also been developed [94, 95, 96]. Indeed, in [95], a block dependent dictionary is used by linking together disparities yielding similar compensation. In [96], directional prediction model combined with linear predictive scheme is proposed for efficient disparity compensation. For the same purpose, the authors proposed in [97] to use the neighborhood of the homologous pixel in the reference image to predict the pixel of the target image and compute the residual image. This computation step is optimized by minimizing the ℓ_1-norm of the resulting prediction error. Note that the disparity is generally encoded using DPCM

(Differencial Pulse Code Modulation) technique followed by an entropy coder while the reference and residual images are often encoded in the transform domain. To this end, the second category of the existing stereo image compression methods have been devoted to the design of efficient decomposition (i.e transform) for coding the reference and residual images. More precisely, some methods have been developed based on the Discrete Cosine Transform (DCT) [99, 100]. However, it has been shown in [101] that residual images contain very narrow vertical edges and DCT yields a moderate energy packing of such images. For this reason, it has been proposed to use a directional DCT to better exploit the specific characteristics of the residual images [100, 102]. Other encoding methods based on wavelet transforms have also been developed in order to provide high quality scalability and progressive reconstruction of the stereo images [103, 104, 105]. Indeed, a family of wavelet-based coders is investigated in [103]. In [104], a coding method based on adaptive lifting scheme has been developed. In this scheme, an adaptive prediction step is performed according to the local gradient information of the reference image. However, the main drawback of this adaptive coding strategy is that it depends on the reference image which has poor quality at low bitrate and results in a significant negative impact on the performance of the stereo image reconstruction process. In [105], a bandelet transform [106] is firstly applied to the left and right images to estimate the disparity map and generate the residual image. Then, the disparity map as well as the bandelet coefficients of the left and residual images are encoded. The main limitation of this method is that it requires to transmit a side information related to the size of each block transform which will affect its performance at low bitrate. In addition to this kind of methods based on the coding of reference and residual images, an alternative approach that does not directly generate a residual image has been proposed in [107] and [108] for grayscale and color stereo images, respectively. More precisely, the approach consists in using a multiscale decomposition based on the concept of vector lifting scheme (VLS). Note that, unlike conventional lifting scheme, the VLS is a joint wavelet decomposition that aims at exploiting the inter-view correlations to generate two compact multi-resolution representations of the left and right images. While a separable decomposition has been carried out in [107], its extended non separable version (NS-VLS) has been developed in [109]. Such extension presents two main advantages. First, it allows to better capture the two dimensional characteristics of the edges which are neither horizontal nor vertical. Moreover, it offers more flexibility in the design of an adaptive transform well adapted to the contents of the input images [110, 111].

3.2.2 Main idea behind the proposed method

To encode the stereo images, we propose to retain the previous NS-VLS decomposition and focus on the built of *content-adaptive decomposition* through sparse optimization algorithms. This is achieved using different ℓ_1 based minimization techniques. It is important to note here that sparse optimization algorithms have been recently employed for still image coding [112]; whereas this work consists in extending them to the context of stereo image coding. While *only* a weighted ℓ_1 minimization technique has already been investigated for stereo image coding [109] and hologram compression [113], this work aims at developing and studying *various* optimization strategies for the design of all the involved filters used with the left and right images. More precisely, in addition to the basic ℓ_2 and ℓ_1 optimization approaches which can be separately applied to each filter of each view, two optimization algorithms based on the weighted ℓ_1 minimization technique are considered. In the first one, we resort to a *full* optimization algorithm where the filters of each view are optimized independently of those used with the other view. However, in the second one, a *joint* optimization algorithm, based on a hybrid weighted ℓ_1 minimization technique, is developed to take into account the inter-view redundancies.

3.3 Non separable vector lifting scheme

Let us first recall that a conventional separable lifting scheme (LS) [114] consists in splitting the input 1D signal into two sets formed by the even and odd samples, respectively. Then, prediction and update steps are applied to generate the detail and approximation signals. Such structure is referred to as P-U (Predict-Update) LS like the 5/3 transform retained in the JPEG2000 coding standard [115]. As shown in [111], a 1D P-U LS has an equivalent 2D non separable structure that can be obtained by splitting the input image into four polyphase components and applying three prediction steps followed by the update one (P-P-P-U structure) to generate three detail subbands and one approximation subband. Based on this observation, the 2D NS-VLS decomposition has been derived [109] where intra prediction steps are performed on the reference (i.e left) image and hybrid prediction steps are employed with the target (i.e right) image to exploit the intra and inter-view redundancies. The main concepts behind this decomposition will be described in what follows.

3.3.1 Analysis structure

The analysis structure of the NS-VLS decomposition is illustrated in Fig. 3.1. While 2D non separable lifting operators are used, it is worth pointing out that the main feature of this VLS-based decomposition concerns the prediction stages. For instance, a conventional P-P-P-U lifting structure is first applied to the left and right images. Since the left image is selected as a reference image and encoded in intra-mode, a hybrid prediction stage is added to the first lifting steps used with the right view to exploit simultaneously the intra and inter-view redundancies based on the information coming from the left view (highlighted with the red color in Fig. 3.1). In addition to the illustration of the basic concept of a NS-VLS, we should note that the main notations have also been included in the above figure to better understand the core mathematical aspects of the proposed optimization algorithms.

Let us now define the different lifting operators used in this structure to generate the wavelet coefficients of the left and right images.

3.3.2 Wavelet representations of the stereo pairs

As a multiscale transform, the decomposition is described for a given resolution level $j \in \mathbb{N}^*$. Let us denote by $I_j^{(l)}$ and $I_j^{(r)}$ the approximation subbands of the left and right images. Note that $j = 0$ corresponds to the original stereo images $I^{(l)}$ and $I^{(r)}$. Moreover, for each view $v \in \{l, r\}$, the image $I_j^{(v)}$ has four polyphase components:

$$\begin{cases} I_{0,j}^{(v)}(m,n) = I_j^{(v)}(2m,2n) \\ I_{1,j}^{(v)}(m,n) = I_j^{(v)}(2m,2n+1), \\ I_{2,j}^{(v)}(m,n) = I_j^{(v)}(2m+1,2n), \\ I_{3,j}^{(v)}(m,n) = I_j^{(v)}(2m+1,2n+1) \end{cases} \tag{3.1}$$

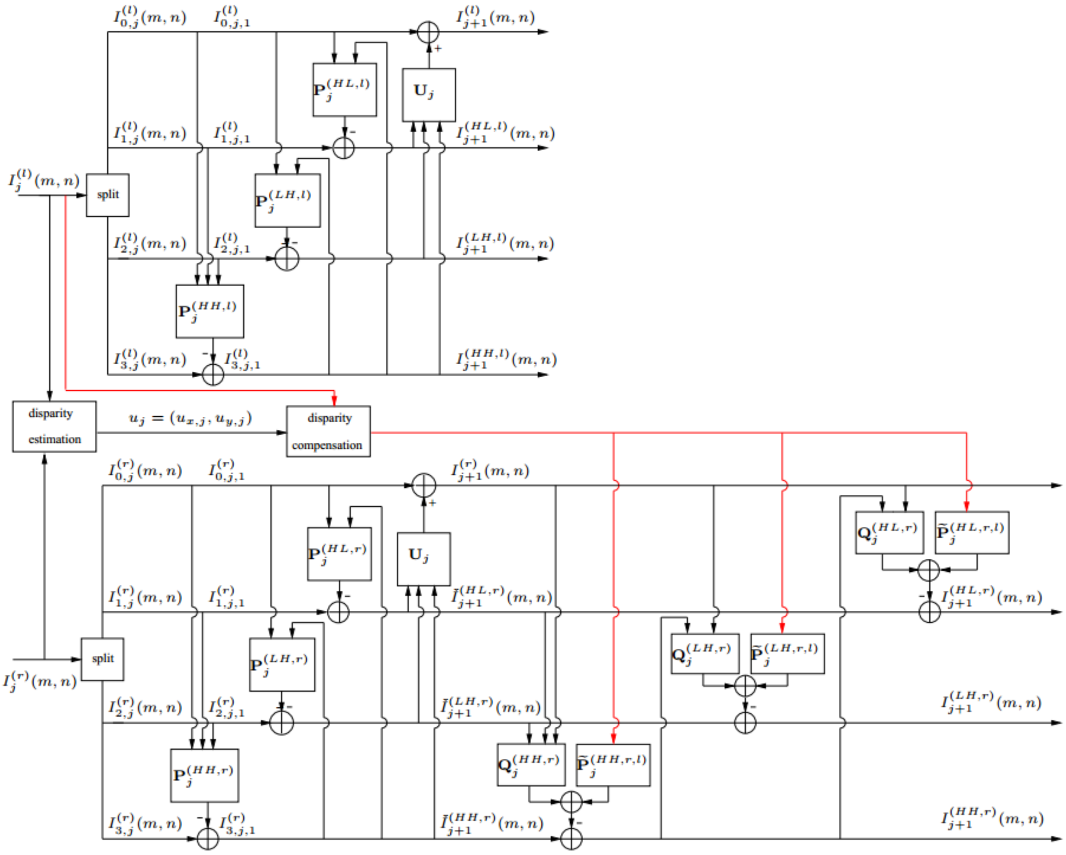


Figure 3.1: *NS-VLS decomposition structure.*

As it can be shown in Fig. 3.1, a non separable lifting stage, composed of three prediction steps and an update one, is applied to the left image to produce three detail subband coefficients oriented diagonally $I_{j+1}^{(HH,l)}$, vertically $I_{j+1}^{(LH,l)}$ and horizontally $I_{j+1}^{(HL,l)}$ as well as the approximation coefficients $I_{j+1}^{(l)}$. These signals can be computed as follows:

$$\begin{aligned} I_{j+1}^{(HH,l)}(m,n) = I_{3,j}^{(l)}(m,n) - \Big(& (\mathbf{P}_{0,j}^{(HH,l)})^\top \mathbf{I}_{0,j}^{(HH,l)} + (\mathbf{P}_{1,j}^{(HH,l)})^\top \mathbf{I}_{1,j}^{(HH,l)} \\ & + (\mathbf{P}_{2,j}^{(HH,l)})^\top \mathbf{I}_{2,j}^{(HH,l)} \Big), \end{aligned} \tag{3.2}$$

$$I_{j+1}^{(LH,l)}(m,n) = I_{2,j}^{(l)}(m,n) - \Big((\mathbf{P}_{0,j}^{(LH,l)})^\top \mathbf{I}_{0,j}^{(LH,l)} + (\mathbf{P}_{1,j}^{(LH,l)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,l)} \Big), \tag{3.3}$$

$$I_{j+1}^{(HL,l)}(m,n) = I_{1,j}^{(l)}(m,n) - \Big((\mathbf{P}_{0,j}^{(HL,l)})^\top \mathbf{I}_{0,j}^{(HL,l)} + (\mathbf{P}_{1,j}^{(HL,l)})^\top \overline{\mathbf{I}}_{j+1}^{(HH,l)} \Big), \tag{3.4}$$

$$\begin{aligned} I_{j+1}^{(l)}(m,n) = I_{0,j}^{(l)}(m,n) + \Big(& (\mathbf{U}_{0,j}^{(HL,l)})^\top \mathbf{I}_{j+1}^{(HL,l)} + (\mathbf{U}_{1,j}^{(LH,l)})^\top \mathbf{I}_{j+1}^{(LH,l)} \\ & + (\mathbf{U}_{2,j}^{(HH,l)})^\top \mathbf{I}_{j+1}^{(HH,l)} \Big), \end{aligned} \tag{3.5}$$

where for each $i \in \{0,1,2\}$ and $o \in \{HL, LH, HH\}$,

- $\mathbf{P}_{i,j}^{(o,l)} = (p_{i,j}^{(o,l)}(s,t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,l)}}$ is the vector of prediction filter coefficients and $\mathcal{P}_{i,j}^{(o,l)}$ denotes its support,

- $\mathbf{I}_{i,j}^{(o,l)} = (I_{i,j}^{(l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{i,j}^{(o,l)}}$ is a reference vector that allows to compute $I_{j+1}^{(o,l)}(m,n)$,

- $\underline{\mathbf{I}}_{j+1}^{(HH,l)} = (I_{j+1}^{(HH,l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{1,j}^{(LH,l)}}$ and $\overline{\mathbf{I}}_{j+1}^{(HH,l)} = (I_{j+1}^{(HH,l)}(m+s,n+t))_{(s,t)\in\mathcal{P}_{1,j}^{(HL,l)}}$ are used in the second and third prediction steps,

- $\mathbf{U}_{i,j}^{(o,l)} = (u_{i,j}^{(o,l)}(s,t))_{(s,t)\in\mathcal{U}_{i,j}^{(o,l)}}$ is an update weighting vector with support $\mathcal{U}_{i,j}^{(o,l)}$,

- $\mathbf{I}_{j+1}^{(o,l)} = (I_{j+1}^{(o,l)}(m+s,n+t))_{(s,t)\in\mathcal{U}_{i,j}^{(o,l)}}$ is the reference vector containing the samples used in the update step.

Unlike the conventional lifting scheme applied to the reference image, an improved one is applied to the target one (i.e the right image). Indeed, let us recall that the key idea behind vector lifting scheme [107] consists in using hybrid (intra and inter) prediction steps. For instance, as it can be seen from Fig. 3.1, the prediction steps used with the target image use some samples from the current view as well as their matching ones in the reference image. To this respect, for the right image, Eqs. (3.2)-(3.5) are firstly applied to produce three intermediate detail subbands and an approximation one denoted respectively by $\check{I}_{j+1}^{(HH,r)}$, $\check{I}_{j+1}^{(LH,r)}$, $\check{I}_{j+1}^{(HL,r)}$ and $I_{j+1}^{(r)}$. Then, a second hybrid prediction stage, composed of three steps, is added to exploit at the same time the intra and inter-view redundancies in the stereo images. This is achieved by using the estimated disparity field denoted by $u_j = (u_{x,j}, u_{y,j})$. For the sake of concision, the disparity compensated left image at a given matching sample (m,n), given by $I_j^{(l)}(m + u_{x,j}(m,n), n + u_{y,j}(m,n))$, is simply replaced by $I_j^{(c)}(m,n)$. Let us denote its corresponding four polyphase components by $I_{0,j}^{(c)}(m,n)$, $I_{1,j}^{(c)}(m,n)$, $I_{2,j}^{(c)}(m,n)$ and $I_{3,j}^{(c)}(m,n)$. Therefore, the final detail subbands of the right image are given by:

$$\begin{aligned} I_{j+1}^{(HH,r)}(m,n) = \check{I}_{j+1}^{(HH,r)}(m,n) - \Big(& (\mathbf{Q}_{0,j}^{(HH,r)})^\top \check{\mathbf{I}}_{0,j+1}^{(HH,r)} + (\mathbf{Q}_{1,j}^{(HH,r)})^\top \check{\mathbf{I}}_{1,j+1}^{(HH,r)} \\ & + (\mathbf{Q}_{2,j}^{(HH,r)})^\top \check{\mathbf{I}}_{2,j+1}^{(HH,r)} + (\widetilde{\mathbf{P}}_{0,j}^{(HH,r,l)})^\top \mathbf{I}_{0,j}^{(HH,c)} + (\widetilde{\mathbf{P}}_{1,j}^{(HH,r,l)})^\top \mathbf{I}_{1,j}^{(HH,c)} \\ & + (\widetilde{\mathbf{P}}_{2,j}^{(HH,r,l)})^\top \mathbf{I}_{2,j}^{(HH,c)} + (\widetilde{\mathbf{P}}_{3,j}^{(HH,r,l)})^\top \mathbf{I}_{3,j}^{(HH,c)} \Big), \end{aligned} \tag{3.6}$$

$$\begin{aligned} I_{j+1}^{(LH,r)}(m,n) = \check{I}_{j+1}^{(LH,r)}(m,n) - \Big(& (\mathbf{Q}_{0,j}^{(LH,r)})^\top \check{\mathbf{I}}_{0,j+1}^{(LH,r)} + (\mathbf{Q}_{1,j}^{(LH,r)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,r)} \\ & + (\widetilde{\mathbf{P}}_{0,j}^{(LH,r,l)})^\top \mathbf{I}_{0,j}^{(LH,c)} + (\widetilde{\mathbf{P}}_{1,j}^{(LH,r,l)})^\top \mathbf{I}_{1,j}^{(LH,c)} + (\widetilde{\mathbf{P}}_{2,j}^{(LH,r,l)})^\top \mathbf{I}_{2,j}^{(LH,c)} \\ & + (\widetilde{\mathbf{P}}_{3,j}^{(LH,r,l)})^\top \mathbf{I}_{3,j}^{(LH,c)} \Big), \end{aligned} \tag{3.7}$$

$$\begin{aligned} I_{j+1}^{(HL,r)}(m,n) = \check{I}_{j+1}^{(HL,r)}(m,n) - \Big(& (\mathbf{Q}_{0,j}^{(HL,r)})^\top \check{\mathbf{I}}_{0,j+1}^{(HL,r)} + (\mathbf{Q}_{1,j}^{(HL,r)})^\top \underline{\mathbf{I}}_{j+1}^{(HH,r)} \\ & + (\widetilde{\mathbf{P}}_{0,j}^{(HL,r,l)})^\top \mathbf{I}_{0,j}^{(HL,c)} + (\widetilde{\mathbf{P}}_{1,j}^{(HL,r,l)})^\top \mathbf{I}_{1,j}^{(HL,c)} + (\widetilde{\mathbf{P}}_{2,j}^{(HL,r,l)})^\top \mathbf{I}_{2,j}^{(HL,c)} \\ & + (\widetilde{\mathbf{P}}_{3,j}^{(HL,r,l)})^\top \mathbf{I}_{3,j}^{(HL,c)} \Big), \end{aligned} \tag{3.8}$$

where for every $i \in \{0,1,2,3\}$ and $o \in \{HL, LH, HH\}$,

- $\mathbf{Q}_{i,j}^{(o,r)} = (q_{i,j}^{(o,r)}(s,t))_{(s,t)\in\mathcal{Q}_{i,j}^{(o,r)}}$ is an intra prediction weighting vector whose support is denoted by $\mathcal{Q}_{i,j}^{(o,r)}$,

- $\widetilde{\mathbf{P}}_{i,j}^{(o,r,l)} = (p_{i,j}^{(o,r,l)}(s,t))_{(s,t)\in\widetilde{\mathcal{P}}_{i,j}^{(o,r,l)}}$ is an inter prediction weighting vector whose support is denoted by $\widetilde{\mathcal{P}}_{i,j}^{(o,r,l)}$,

- $\check{\mathbf{I}}_{0,j+1}^{(o,r)} = (I_{j+1}^{(r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{0,j}^{(o,r)}}$ is a reference vector used to compute $I_{j+1}^{(o,r)}(m,n)$,

- $\check{\mathbf{I}}_{1,j+1}^{(HH,r)} = (\check{I}_{j+1}^{(HL,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(HH,r)}}$ and $\check{\mathbf{I}}_{2,j+1}^{(HH,r)} = (\check{I}_{j+1}^{(LH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{2,j}^{(HH,r)}}$ are two reference vectors used to compute $I_{j+1}^{(HH,r)}(m,n)$,

- $\underline{\mathbf{I}}_{j+1}^{(HH,r)} = (I_{j+1}^{(HH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(LH,r)}}$ and $\bar{\mathbf{I}}_{j+1}^{(HH,r)} = (I_{j+1}^{(HH,r)}(m+s,n+t))_{(s,t)\in\mathcal{Q}_{1,j}^{(HL,r)}}$ are two intra prediction vectors used to compute $I_{j+1}^{(LH,r)}(m,n)$ and $I_{j+1}^{(HL,r)}(m,n)$,

- $\mathbf{I}_{i,j}^{(o,c)} = (I_{i,j}^{(c)}(m+s,n+t))_{(s,t)\in\widetilde{\mathcal{P}}_{i,j}^{(o,r,l)}}$ is a reference vector containing the matching samples used to compute $I_{j+1}^{(o,r)}(m,n)$.

Finally, at the last resolution level $j = J$, instead of coding the approximation subband $I_J^{(r)}$, a residual subband $e_J^{(r)}$ is generated by computing the difference between the right approximation subband and the disparity compensated left one:

$$e_J^{(r)}(m,n) = I_J^{(r)}(m,n) - I_J^{(c)}(m,n). \tag{3.9}$$

Once the considered NS-VLS has been defined, we investigate in the next section techniques for optimizing sparse criteria to design the lifting operators used with the left and right images.

3.4 Proposed sparse optimization algorithms

Since the coding performance of wavelet-based coding scheme depends on the choice of the lifting operators, a great attention should be paid to the design of the prediction and update filters of both views. To this respect, two kinds of optimization strategies could be adopted and will be described in what follows.

3.4.1 Independent full optimization approach

A straightforward solution consists in applying classical optimization methods used in the context of lifting-based still image coding to each view separately. Thus, the two lifting structures used with the left and right images can be firstly optimized in an independent way. To this end, we will resort to ℓ_2, ℓ_1 and weighted ℓ_1 based minimization techniques. While the update filter of each view $\mathbf{U}_j^{(v)}$, with $v \in \{r, l\}$, is often optimized by minimizing the error between the approximation subband $I_{j+1}^{(v)}$ and the decimated subband obtained after an ideal low-pass filtering of $I_j^{(v)}$ [111], we will focus here on the optimization methods for designing the different prediction filters. To this respect, one should first note that the standard approach consists in minimizing the variance (i.e the ℓ_2-norm) of the output detail subband since the latter can be seen as a prediction error [116, 117]. The main advantage of this approach is its simplicity since the optimal prediction filters are solutions of a linear system of equations.

ℓ_1-based minimization technique

In addition to ℓ_2-minimization technique, ℓ_1-based minimization has been recently investigated in the context of *one* stage lifting structure for still image coding purpose [112]. It is important to note here that the use of ℓ_1 criterion presents two main advantages. First, minimizing an ℓ_1 criterion allows to generate sparse representation which could achieve good coding performance [118]. Moreover, from the information theory point of view, it has been shown that, at high bitrate, the minimization of the entropy of the detail subbands is closely related to the minimization of their ℓ_β-norm where β is the shape parameter of a generalized gaussian distribution (GGD) used for modeling the detail coefficients [119]. Indeed, knowing that the wavelet detail subbands $I_{j+1}^{(o,v)}$ are generally multiplied by the weights $\sqrt{w_{j+1}^{(o,v)}}$ before the entropy encoding, and if we consider $\beta = 1$, the resulting differential entropy can be obtained as follows:

$$\frac{1}{M_j N_j \alpha_{j+1}^{(o,v)} \ln(2)} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \left| I_{j+1}^{(o,v)}(m,n) \right| + \log_2 \left(2\alpha_{j+1}^{(o,v)} \sqrt{w_{j+1}^{(o,v)}} \right) \tag{3.10}$$

where (M_j, N_j) represent the dimensions of the subband $I_{j+1}^{(o,v)}$, $\alpha_{j+1}^{(o,v)}$ is the scale parameter of the GGD which can be estimated using a classical maximum likelihood estimate, and the weights $w_{j+1}^{(o,v)}$ are computed based on the wavelet filters used for the reconstruction process as proposed in [120, 121].

Therefore, instead of minimizing the ℓ_2-norm, each prediction filter $\mathbf{P}_j^{(o,l)}$, $\mathbf{P}_j^{(o,r)}$ and $\mathbf{P}_j^{(o,r,l)} = \left(\mathbf{Q}_j^{(o,r)}, \tilde{\mathbf{P}}_j^{(o,r,l)}\right)^\top$ could be optimized by minimizing the ℓ_1-norm of their output detail subbands $I_{j+1}^{(o,l)}$ and $I_{j+1}^{(o,r)}$. Thus, for the intra prediction filters $\mathbf{P}_j^{(o,l)}$ and $\mathbf{P}_j^{(o,r)}$, the criterion is expressed as:

$$\begin{aligned}
&\forall\ v \in \{r,l\}, \forall\ o \in \{HL, LH, HH\}, \forall\ i \in \{1,2,3\},\\
\mathcal{J}_{\ell_1}^{(v)}(\mathbf{P}_j^{(o,v)}) &= \sum_{m=1}^{M_j}\sum_{n=1}^{N_j} \left| I_{j+1}^{(o,v)}(m,n) \right| \\
&= \sum_{m=1}^{M_j}\sum_{n=1}^{N_j} \left| I_{i,j}^{(v)}(m,n) - (\mathbf{P}_j^{(o,v)})^\top \tilde{\mathbf{I}}_j^{(o,v)}(m,n) \right| \qquad (3.11)
\end{aligned}$$

with $I_{i,j}^{(v)}(m,n)$ is the $(i+1)^{th}$ polyphase component of the view $I_j^{(v)}$ to be predicted, $\mathbf{P}_j^{(o,v)}$ is the prediction operator vector to be optimized, and $\tilde{\mathbf{I}}_j^{(o,v)}(m,n)$ is the reference vector containing the samples used in the prediction step. Similarly, for the hybrid prediction filters $\mathbf{P}_j^{(o,r,l)}$ used in the second lifting stage with the right image, the ℓ_1 criterion will be rewritten as:

$$\begin{aligned}
&\forall\ o \in \{HL, LH, HH\}, \forall\ i \in \{1,2,3\},\\
\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(o,r,l)}) &= \sum_{m=1}^{M_j}\sum_{n=1}^{N_j} \left| \check{I}_{i,j}^{(r)}(m,n) - (\mathbf{P}_j^{(o,r,l)})^\top \tilde{\mathbf{I}}_j^{(o,r,l)}(m,n) \right| \qquad (3.12)
\end{aligned}$$

where $\tilde{\mathbf{I}}_j^{(o,r,l)}(m,n)$ is a reference vector containing the samples from right and disparity compensated left images used in the prediction step, and $\check{I}_{i,j}^{(r)}$ is the polyphase component to be predicted in the second lifting stage. According to

Fig. 3.1, the four polyphase components of the second lifting stage are defined as:

$$\begin{cases} \check{I}^{(r)}_{0,j}(m,n) = I^{(r)}_{j+1}(m,n) \\ \check{I}^{(r)}_{1,j}(m,n) = \check{I}^{(HL,r)}_{j+1}(m,n) \\ \check{I}^{(r)}_{2,j}(m,n) = \check{I}^{(LH,r)}_{j+1}(m,n) \\ \check{I}^{(r)}_{3,j}(m,n) = \check{I}^{(HH,r)}_{j+1}(m,n) \end{cases} \tag{3.13}$$

To minimize this criterion, we propose to use the proximity operators tool [122] which has been found to be efficient for solving nonsmooth optimization problem [123, 124]. Based on this tool, the minimization of the above ℓ_1 criterion (for example the criterion given by Eq. (3.11)) is equivalent to the following minimization problem:

$$\forall\ o \in \{HL, LH, HH\}, \forall\ i \in \{1,2,3\},$$
$$\min_{\mathbf{z}^{(o,v)}_j \in V} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \left| I^{(v)}_{i,j}(m,n) - z^{(o,v)}_j(m,n) \right| + \imath_V(\mathbf{z}^{(o,v)}_j), \tag{3.14}$$

where $\imath_V$ is the indicator function and V is the vector space given by:

$$V = \{\mathbf{z}^{(o,v)}_j = \left(z^{(o,v)}_j(m,n)\right)_{\substack{1\le m\le M_j \\ 1\le n\le N_j}} \in \mathbb{R}^{M_j \times N_j} \mid \exists\ \mathbf{P}^{(o,v)}_j,$$
$$\forall\ (m,n) \in \{1,\dots,M_j\} \times \{1,\dots,N_j\}, z^{(o,v)}_j(m,n) = (\mathbf{P}^{(o,v)}_j)^\top \tilde{\mathbf{I}}^{(o,v)}_j(m,n)\}.$$

After that, the Douglas Rachford (DR) algorithm will be applied to solve our minimization problem and obtain the optimized prediction filter $\mathbf{P}^{(o,v)}_j$. For more details on the proximity operators as well as the DR algorithm, the reader is referred to [112].

Weighted ℓ_1-based minimization technique

In the previous part, each prediction filter $\mathbf{P}^{(o,v)}_j$ has been separately optimized by minimizing the ℓ_1-norm of its associated detail subband $I^{(o,v)}_{j+1}$. However, if we focus on the lifting structure applied to the left image, it can be seen in Fig. 3.1 that the diagonal detail coefficients resulting from the first prediction step are used in the second and third prediction steps to generate the left detail coefficients oriented vertically $I_{j+1}^{(LH,l)}$ and horizontally $I_{j+1}^{(HL,l)}$, respectively. Therefore, instead of minimizing the ℓ_1-norm of the diagonal detail coefficients, it becomes more interesting to optimize the first prediction filter $\mathbf{P}_j^{(HH,l)}$ by minimizing a weighted sum of the ℓ_1-norms of the three detail subbands of the left image. Concerning the filters $\mathbf{P}_j^{(LH,l)}$ and $\mathbf{P}_j^{(HL,l)}$, they will be simply optimized by minimizing the ℓ_1-norm of their corresponding detail coefficients $I_{j+1}^{(LH,l)}$ and $I_{j+1}^{(HL,l)}$ since the second and third predictions are two independent steps.

Regarding the prediction filters used with the lifting structure applied to the right image, it can be also seen from Fig. 3.1 that the first three intermediate detail coefficients $\check{I}_{j+1}^{(o,r)}$ as well as the final diagonal detail ones $I_{j+1}^{(HH,r)}$ are involved in the last two prediction steps to generate the final horizontal and vertical detail subbands. Therefore, and similarly to the left image, the first four prediction filters, $\mathbf{P}_j^{(o,r)}$ with $o \in \{HH, LH, HL\}$ and $\mathbf{P}_j^{(HH,r,l)}$, used with the right image should be optimized by minimizing the weighted sum of the ℓ_1-norms of the three detail subbands of the right image. Finally, the last prediction filters $\mathbf{P}_j^{(LH,r,l)}$ and $\mathbf{P}_j^{(HL,r,l)}$ are optimized by minimizing the ℓ_1-norm of their corresponding detail coefficients $I_{j+1}^{(LH,r)}$ and $I_{j+1}^{(HL,r)}$.

Therefore, the weighted ℓ_1 criterion used with the left and right images can be expressed for each view as follows:

$$\begin{aligned}
&\forall\ v \in \{r,l\}, \forall\ o \in \{HL, LH, HH\},\\
&\mathcal{J}_{w\ell_1}^{(v)}(\mathbf{P}_j^{(o,v)}) = \sum_{o \in \{HL,LH,HH\}} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \frac{1}{\alpha_{j+1}^{(o,v)}} \left| I_{j+1}^{(o,v)}(m,n) \right|.
\end{aligned} \tag{3.15}$$

To minimize this weighted ℓ_1 criterion, we need first to rewrite the weighted criterion (i.e $I_{j+1}^{(o,v)}$) as a function of the filter to be optimized $\mathbf{P}_j^{(o,v)}$. To this end, and since the weighted ℓ_1 minimization concerns only the first prediction filter for the left image and the first four prediction filters for the right image, let us introduce the notation $\left(I_{i,j,q}^{(v)}\right)_{i \in \{0,1,2,3\}}$ the four polyphase components obtained from the inputs $\left(\hat{I}_{i,j,q-1}^{(v)}\right)_{i \in \{0,1,2,3\}}$ after the q-th prediction step while $q \in \{1\}$

(resp. $q \in \{1, 2, 3, 4\}$) in the case of the left (resp. right) view. To illustrate these components, the terms $\left(I^{(v)}_{i,j,1}\right)_{i\in\{0,1,2,3\}}$ (i.e for $q = 1$) have been inserted in Fig. 3.1 after the first prediction step used with both views. Thus, for each $i \in \{0, 1, 2, 3\}$, we have:

$$\begin{cases} \hat{I}^{(v)}_{i,j,q-1}(m,n) = I^{(v)}_{i,j}(m,n) & \text{for} \quad q = 1, \quad \forall\, v \in \{r, l\} \\ \hat{I}^{(r)}_{i,j,q-1}(m,n) = I^{(r)}_{i,j,q-1}(m,n) & \text{for} \quad q \in \{2,3\} \\ \hat{I}^{(r)}_{i,j,q-1}(m,n) = \check{I}^{(r)}_{i,j}(m,n) & \text{for} \quad q = 4 \end{cases} \tag{3.16}$$

Based on these notations, each detail subband $I^{(o,v)}_{j+1}$ can be written as a function of the filter to be optimized $\mathbf{P}^{(o,v)}_{j}$ as follows:

$$\begin{aligned} &\forall\, o \in \{HH, LH, HL\}, \\ &I^{(o,v)}_{j+1}(m,n) = y^{(o,v)}_{j,q}(m,n) - (\mathbf{P}^{(o,v)}_{j})^{\top}\mathbf{I}^{(o,v)}_{j,q}(m,n) \end{aligned} \tag{3.17}$$

where the signal to be predicted $y^{(o,v)}_{j,q}(m,n)$ as well as the reference vector $\mathbf{I}^{(o,v)}_{j,q}(m,n)$ are given by:

$$y^{(o,v)}_{j,q}(m,n) = \sum_{i'\in\mathbb{I}_i}\sum_{k,l} h^{(o,v)}_{i',j,q}(k,l) I^{(v)}_{i',j,q}(m-k, n-l) + \sum_{k,l} h^{(o,v)}_{i,j,q}(k,l)\hat{I}^{(v)}_{i,j,q-1}(m-k, n-l), \tag{3.18}$$

$$\mathbf{I}^{(o,v)}_{j,q}(m,n) = \left(\sum_{k,l} h^{(o,v)}_{i,j,q}(k,l)\hat{I}^{(v)}_{i',j,q-1}(m-k-r, n-l-s)\right)_{\substack{(r,s)\in\mathcal{P}^{(o,v)}_{j} \\ i'\in\mathbb{I}_i}} \tag{3.19}$$

with $\mathbb{I}_i = \{0, 1, 2, 3\} \backslash \{i\}$ and i is the index number of the polyphase component to be predicted by the current prediction filter under optimization.

Once the different terms involved in the weighted ℓ_1 criterion are defined, the DR algorithm in a three-fold product space could be applied to solve our minimization problem. More details about DR algorithm through a formulation in three fold product space can be found in [112] and references therein.

Full optimization algorithm

As mentioned before, for the left image, the optimization of the filter $\mathbf{P}_j^{(HH,l)}$ depends on the optimization of the filters $\mathbf{P}_j^{(LH,l)}$ and $\mathbf{P}_j^{(HL,l)}$ since the weighted sum of the ℓ_1-norms of all the detail subband coefficients is minimized. On the other hand, the optimization of the filters $\mathbf{P}_j^{(LH,l)}$ and $\mathbf{P}_j^{(HL,l)}$ depends on the optimization of $\mathbf{P}_j^{(HH,l)}$ since the latter allows to compute the diagonal detail subband which is used in the second and third prediction steps. Similarly, for the right image, the optimization of the first four prediction filters depends on the optimization of the last two ones, and vice-versa. Therefore it becomes interesting to resort to an iterative algorithm which jointly optimizes the different prediction filters. To this respect, we start by optimizing all the filters used with the left image independently of the right one. Then, all the filters of the right image will be optimized. Thus, our first independent full optimization algorithm can be described as follows:

Algorithm 1: Independent full optimization algorithm

① Optimization of the left image filters:

(a) Initialize the iteration number it to 0

- Optimize separately the three prediction filters $\mathbf{P}^{(o,l)}$ by minimizing the ℓ_1 criterion $\mathcal{J}_{\ell_1}^{(l)}(\mathbf{P}_j^{(o,l)})$. The new filters will be designated respectively by $\mathbf{P}_j^{(HH,l,0)}$, $\mathbf{P}_j^{(LH,l,0)}$, and $\mathbf{P}_j^{(HL,l,0)}$.
- Optimize the update filter [111].
- Compute the constant values $\frac{1}{\alpha_{j+1}^{(o,l,0)}}$, the weights $w_{j+1}^{(o,l,0)}$ and the differential entropy of the three resulting detail subbands.

(b) for $it = 1, 2, 3, \ldots$

- Set $\mathbf{P}_j^{(LH,l)} = \mathbf{P}_j^{(LH,l,it-1)}$, $\mathbf{P}_j^{(HL,l)} = \mathbf{P}_j^{(HL,l,it-1)}$, and optimize $\mathbf{P}_j^{(HH,l)}$ by minimizing the weighted ℓ_1 criterion $\mathcal{J}_{w\ell_1}^{(l)}(\mathbf{P}_j^{(HH,l)})$. Let $\mathbf{P}_j^{(HH,l,it)}$ be the new optimal filter.
- Set $\mathbf{P}_j^{(HH,l)} = \mathbf{P}_j^{(HH,l,it)}$, and optimize $\mathbf{P}_j^{(LH,l)}$ as well as $\mathbf{P}_j^{(HL,l)}$ by minimizing $\mathcal{J}_{\ell_1}^{(l)}(\mathbf{P}_j^{(LH,l)})$ and $\mathcal{J}_{\ell_1}^{(l)}(\mathbf{P}_j^{(HL,l)})$, respectively. Let $\mathbf{P}_j^{(LH,l,it)}$ and $\mathbf{P}_j^{(HL,l,it)}$ be the new optimal filters.

• Optimize the update filter.

• Compute the new constant values $\frac{1}{\alpha_{j+1}^{(o,l,it)}}$, the weights $w_{j+1}^{(o,l,it)}$ and the differential entropy of the three resulting detail subbands.

② Optimization of the right image filters:

(a) Initialize the iteration number it to 0

• Optimize separately the three intra prediction filters $\mathbf{P}_j^{(o,r)}$ by minimizing the ℓ_1 criterion $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(o,r)})$.

• Optimize the update filter.

• Optimize separately the three inter prediction filters $\mathbf{P}_j^{(o,r,l)}$ by minimizing the ℓ_1 criterion $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(o,r,l)})$.

• Compute the constant values $\frac{1}{\alpha_{j+1}^{(o,r,0)}}$, the weights $w_{j+1}^{(o,r,0)}$ and the differential entropy of the three resulting detail subbands.

(b) for $it = 1, 2, 3, \ldots$

• Optimize $\mathbf{P}_j^{(HH,r)}$, while setting all the other filters equal to those obtained in the previous iteration $(it-1)$, by minimizing the weighted criterion $\mathcal{J}_{w\ell_1}^{(r)}(\mathbf{P}_j^{(HH,r)})$. Let $\mathbf{P}_j^{(HH,r,it)}$ be the new optimal filter.

• Set $\mathbf{P}_j^{(HH,r)} = \mathbf{P}_j^{(HH,r,it)}$, and optimize $\mathbf{P}_j^{(LH,r)}$ by minimizing the weighted criterion $\mathcal{J}_{w\ell_1}^{(r)}(\mathbf{P}_j^{(LH,r)})$. Let $\mathbf{P}_j^{(LH,r,it)}$ be the new optimal filter.

• Set $\mathbf{P}_j^{(LH,r)} = \mathbf{P}_j^{(LH,r,it)}$, and optimize $\mathbf{P}_j^{(HL,r)}$ by minimizing the weighted criterion $\mathcal{J}_{w\ell_1}^{(r)}(\mathbf{P}_j^{(HL,r)})$. Let $\mathbf{P}_j^{(HL,r,it)}$ be the new optimal filter.

• Optimize the update filter.

• Set $\mathbf{P}_j^{(HL,r)} = \mathbf{P}_j^{(HL,r,it)}$, and optimize $\mathbf{P}_j^{(HH,r,l)}$ by minimizing the weighted criterion $\mathcal{J}_{w\ell_1}^{(r)}(\mathbf{P}_j^{(HH,r,l)})$. Let $\mathbf{P}_j^{(HH,r,l,it)}$ be the new optimal filter.

• Set $\mathbf{P}_j^{(HH,r,l)} = \mathbf{P}_j^{(HH,r,l,it)}$, and optimize $\mathbf{P}_j^{(LH,r,l)}$ as well as $\mathbf{P}_j^{(HL,r,l)}$ by minimizing $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(LH,r,l)})$ and $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(HL,r,l)})$, respectively. Let $\mathbf{P}_j^{(LH,r,l,it)}$ and $\mathbf{P}_j^{(HL,r,l,it)}$ be the new optimal filters.

• Compute the new constant values $\frac{1}{\alpha_{j+1}^{(o,r,it)}}$, the weights $w_{j+1}^{(o,r,it)}$ and the differential entropy of the three resulting detail subbands.

3.4.2 Joint optimization approach

According to Fig. 3.1, the lifting stage applied to the left image is similar to the first stage applied to the right image used to generate the approximation subband $I_{j+1}^{(r)}$ and three intermediate detail subbands $\check{I}_{j+1}^{(o,r)}$ with $o \in \{HH, LH, HL\}$. Moreover, one of the main characteristics of stereo images is that they present high inter-view correlations since they correspond to the same 3D scene. Therefore, instead of optimizing each view independently of the other one, it becomes interesting to design a joint optimization approach to take into account the previous observations.

Hybrid weighted ℓ_1 minimization technique

To exploit the correlation existing between the left and right, we propose first to assume that the filters of the first lifting stage employed with the right image $(\mathbf{P}_j^{(o,r)}, \mathbf{U}_j^{(r)})$ are similar to those used with the left image $(\mathbf{P}_j^{(o,l)}, \mathbf{U}_j^{(l)})$. Thus, for the sake of concision, the three intra prediction filters as well as the update one are simply denoted by $\mathbf{P}_j^{(o)}$ and $\mathbf{U}_j$:

$$\begin{cases} \mathbf{P}_j^{(HH,r)} = \mathbf{P}_j^{(HH,l)} = \mathbf{P}_j^{(HH)}, \\ \mathbf{P}_j^{(LH,r)} = \mathbf{P}_j^{(LH,l)} = \mathbf{P}_j^{(LH)}, \\ \mathbf{P}_j^{(HL,r)} = \mathbf{P}_j^{(HL,l)} = \mathbf{P}_j^{(HL)}, \\ \mathbf{U}_j^{(r)} = \mathbf{U}_j^{(l)} = \mathbf{U}_j. \end{cases} \tag{3.20}$$

Moreover, since these filters are applied both to the left and right images, we also propose to design a hybrid weighted ℓ_1 criterion defined simultaneously on the stereo pairs. More precisely, this criterion is the weighted sum of the ℓ_1-norm of the three detail subbands of the left image as well as the three intermediate detail subbands of the right one. Therefore, the new hybrid weighted ℓ_1 criterion can be expressed as follows:

$$\forall\ o \in \{HL, LH, HH\},$$
$$\mathcal{J}^{(r,l)}_{w\ell_1}(\mathbf{P}^{(o)}_j) = \sum_{o\in\{HL,LH,HH\}} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \Big(\frac{1}{\alpha^{(o,l)}_{j+1}} \Big| I^{(o,l)}_{j+1}(m,n) \Big| + \frac{1}{\alpha^{(o,r)}_{j+1}} \Big| \check{I}^{(o,r)}_{j+1}(m,n) \Big| \Big) \tag{3.21}$$

Similarly to the weighted ℓ_1 criterion given by Eq. (3.15), the new hybrid one will also be minimized using the DR algorithm in a product space.

Once the prediction filters used in the first lifting stage of both views have been jointly optimized, the last three inter prediction filters used with the right image are optimized as performed in the previous optimization algorithm (i.e by minimizing the weighted sum of the ℓ_1-norms of the three detail subbands of the right image).

Therefore, the second optimization algorithm can be summarized as follows.

Algorithm 2: Joint optimization algorithm

① Optimization of the intra prediction filters used with the left and right images:

(a) Initialize the iteration number it to 0

- Optimize separately the three prediction filters $\mathbf{P}^{(o)}_j$ by minimizing the ℓ_1 criterion $\mathcal{J}^{(l)}_{\ell_1}(\mathbf{P}^{(o)}_j)$. The new filters will be designated respectively by $\mathbf{P}^{(HH,0)}_j$, $\mathbf{P}^{(LH,0)}_j$, and $\mathbf{P}^{(HL,0)}_j$.
- Optimize the update filter of the left image.
- Set the intra prediction filters and update one equal to those obtained with the left image (Eq. 3.20).
- Compute the constant values $\frac{1}{\alpha^{(o,l,0)}_{j+1}}$ and $\frac{1}{\alpha^{(o,r,0)}_{j+1}}$, the weights $w^{(o,l,0)}_{j+1}$ and $w^{(o,r,0)}_{j+1}$, as well as the differential entropy of the six resulting detail subbands.

(b) for $it = 1, 2, 3, \dots$

- Set $\mathbf{P}_j^{(LH)} = \mathbf{P}_j^{(LH,it-1)}$, $\mathbf{P}_j^{(HL)} = \mathbf{P}_j^{(HL,it-1)}$, and optimize $\mathbf{P}_j^{(HH)}$ by minimizing the hybrid weighted ℓ_1 criterion $\mathcal{J}_{w\ell_1}^{(r,l)}(\mathbf{P}_j^{(HH)})$. Let $\mathbf{P}_j^{(HH,it)}$ be the new optimal filter.
- Set $\mathbf{P}_j^{(HH)} = \mathbf{P}_j^{(HH,it)}$, and optimize $\mathbf{P}_j^{(LH)}$ as well as $\mathbf{P}_j^{(HL)}$ by minimizing $\mathcal{J}_{w\ell_1}^{(r,l)}(\mathbf{P}_j^{(LH)})$ and $\mathcal{J}_{w\ell_1}^{(r,l)}(\mathbf{P}_j^{(HL)})$, respectively. Let $\mathbf{P}_j^{(LH,it)}$ and $\mathbf{P}_j^{(HL,it)}$ be the new optimal filters.
- Optimize the update filter.
- Compute the new constant values $\frac{1}{\alpha_{j+1}^{(o,l,it)}}$ and $\frac{1}{\alpha_{j+1}^{(o,r,it)}}$, the weights $w_{j+1}^{(o,l,it)}$ and $w_{j+1}^{(o,r,it)}$, as well as the differential entropy of the six resulting detail subbands.

② Optimization of the remaining inter prediction filters used with the right image:

(a) Initialize the iteration number it to 0

- Apply the first intra lifting stage to the right image using the optimal filters obtained with the left image, and optimize separately the three inter prediction filters $\mathbf{P}_j^{(o,r,l)}$ by minimizing the ℓ_1 criterion $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(o,r,l)})$. The new filters will be denoted respectively by $\mathbf{P}_j^{(HH,r,l,0)}$, $\mathbf{P}_j^{(LH,r,l,0)}$, and $\mathbf{P}_j^{(HL,r,l,0)}$.
- Compute the constant values $\frac{1}{\alpha_{j+1}^{(o,r,0)}}$, the weights $w_{j+1}^{(o,r,0)}$ and the differential entropy of the three resulting detail subbands.

(b) for $it = 1, 2, 3, \ldots$

- Set $\mathbf{P}_j^{(LH,r,l)} = \mathbf{P}_j^{(LH,r,l,it-1)}$, $\mathbf{P}_j^{(HL,r,l)} = \mathbf{P}_j^{(HL,r,l,it-1)}$, and optimize $\mathbf{P}_j^{(HH,r,l)}$ by minimizing the weighted criterion $\mathcal{J}_{w\ell_1}^{(r)}(\mathbf{P}_j^{(HH,r,l)})$. Let $\mathbf{P}_j^{(HH,r,l,it)}$ be the new optimal filter.
- Set $\mathbf{P}_j^{(HH,r,l)} = \mathbf{P}_j^{(HH,r,l,it)}$, and optimize $\mathbf{P}_j^{(LH,r,l)}$ as well as $\mathbf{P}_j^{(HL,r,l)}$ by minimizing $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(LH,r,l)})$ and $\mathcal{J}_{\ell_1}^{(r)}(\mathbf{P}_j^{(HL,r,l)})$, respectively. Let $\mathbf{P}_j^{(LH,r,l,it)}$ and $\mathbf{P}_j^{(HL,r,l,it)}$ be the new optimal filters.
- Compute the new constant values $\frac{1}{\alpha_{j+1}^{(o,r,it)}}$, the weights $w_{j+1}^{(o,r,it)}$ and the differential entropy of the three resulting detail subbands.

It is important to note here that the convergence of the two proposed optimization algorithms is achieved in few iterations (after about 5 or 6 iterations) where the weighted ℓ_1 minimization technique performed on each prediction filter takes about 4-5 seconds for an image of size 512×512 using a Matlab implementation and a computer with an Intel Core i7 processor (3.4 GHz). For instance, compared to the optimization strategy developed in [109], the proposed joint optimization algorithm increases slightly the execution time (2 seconds per filter) since a hybrid weighted ℓ_1 minimization technique (given by Eq. (3.21)) is employed. Moreover, this joint algorithm presents two main advantages compared to the first independent full one. Indeed, in addition to an efficient exploitation of the characteristics of the stereo images through the design of a hybrid criterion, it simplifies the optimization process and reduces the bitrate of the filter coefficients that should be transmitted to the decoder (thanks to the assumption given by Eq. (3.20)).

3.5 Experimental results

Simulations were conducted on different stereo images taken from various datasets such as VASC CMU and middlebury ones [125, 126]. In order to illustrate the proposed sparse optimization of NS-VLS in the context of stereo image coding, and since our non separable lifting structure is a 2D extension of 1D P-U LS (as explained at the beginning of Sec. 3.3), we will consider the P-U 5/3 LS, known also as (2,2) wavelet transform, which has been selected for the JPEG2000 coding standard. Note that the spatial prediction and update supports of its extended 2D non separable structure (used with the left image and the right one in the first lifting stage) can be found in [111]. For the second lifting stage used with the right image, the intra prediction filters $\mathcal{Q}_{i,j}^{(o,r)}$ have the same spatial supports of the other prediction filters used with the left image as well as the first intra lifting stage employed with the right image, while the spatial supports of the inter prediction filters $\widetilde{\mathbf{P}}_{i,j}^{(o,r,l)}$ are defined by the set $\widetilde{\mathcal{P}}_{i,j}^{(o,r,l)} = \{(s,t),\ \text{with } s \in \{-1,0,1\} \text{ and } t \in \{-1,0,1\}\}$ (where, $s = t = 0$

corresponds to the matching pixel in the disparity compensated left image $I_j^{(c)}(m,n)$ of the current sample to be predicted $I_{i,j}^{(r)}(m,n))$.
Thus, to show the performance of the proposed optimization methods, we will consider the following ones carried out over three resolution levels:

- The first one consists in coding independently the left and right images by applying the 5/3 transform to each view. In the following, this method will be denoted by "Independent".

- The second method represents the state-of-the-art method which consists in coding the left image and the residual one by using the 5/3 wavelet transform. Let us recall that this approach, which will be designated by "Residual", is behind most of the developed stereo image coding schemes.

- While the residual image is generated in the previous method by computing the prediction error between each pixel of the target image and its corresponding one in the reference one, the third method proposed in [97] aims to use the neighborhood of the homologous pixel to predict the pixel of the target image. This computation step is optimized by minimizing the ℓ_1-norm of the resulting prediction error. This method will be designated by "Residual-OPT-L1 [7]".

- The fourth one corresponds to the NS-VLS where the prediction filters are optimized separately by minimizing the variance (i.e ℓ_2-norm) of the detail coefficients. This method will be denoted by "NS-VLS-OPT-L2".

- The fifth method corresponds to the NS-VLS where the prediction filters are optimized separately by minimizing the ℓ_1-norm of the detail coefficients. This method will be denoted by "NS-VLS-OPT-L1".

- The sixth and seventh methods represent the proposed independent full and joint algorithms used to optimize the NS-VLS. These methods will be designated by "NS-VLS-OPT-WL1-Full" and "NS-VLS-OPT-WL1-Joint", respectively.

All these approaches are firstly compared in terms of rate-distortion performance. Figs. 3.2 and 3.3 illustrate the variations of the PSNR versus the bitrate for the "houseof" and "ball" stereo images. Note that the average bitrate as well as the average reconstruction error have been used to evaluate the performance of all these methods. It can be observed that residual-based coding scheme leads to better results compared to the independent coding scheme especially at low bitrate. The optimized residual scheme [97] outperforms the previous one by about 0.15-0.5 dB. Moreover, using sparse ℓ_1 minimization technique improves the ℓ_2 one by achieving a gain of about 0.2-0.4 dB in terms of PSNR. Further improvements are achieved by resorting to the proposed fully and joint weighted ℓ_1 minimization techniques. The gain is about 0.1-0.2 dB compared to the standard ℓ_1 minimization technique. It should be also noted that the results obtained by the joint optimization approach are close to those obtained with the independent fully approach. For instance, one can observe that a very small improvement is achieved by the independent full optimization strategy. Such behavior is expected since, in the full optimization approach, all filters of both views are optimized, whereas in the joint optimization strategy, the intra prediction filters used in the first lifting stage with the right view are assumed to be equal to those used with the left image. Thus, due to the aforementioned advantages of the joint optimization approach (simplifying the optimization process and reducing the transmission cost of the filter coefficients), the latter optimization algorithm is more appropriate from a practical point of view.

After that, and since the joint and full independent optimization methods have similar performances, we will evaluate the relative gain of the proposed joint optimization algorithm "NS-VLS-OPT-WL1-Joint" by using the Bjontegaard metric [127]. For instance, the gains of the joint optimization algorithm with respect to the standard ℓ_2 one "NS-VLS-OPT-L2" as well as the ℓ_1 one "NS-VLS-OPT-L1" are provided in Tables 3.1 and 3.2 for low, middle and high bitrates, which are obtained by considering the following four bitrate points $\{0.15, 0.2, 0.25, 0.3\}$, $\{0.6, 0.65, 0.7, 0.75\}$ and $\{1.25, 1.3, 1.35, 1.4\}$ bpp, respectively. Note that a bitrate saving with respect to a given reference method corresponds to a negative value.

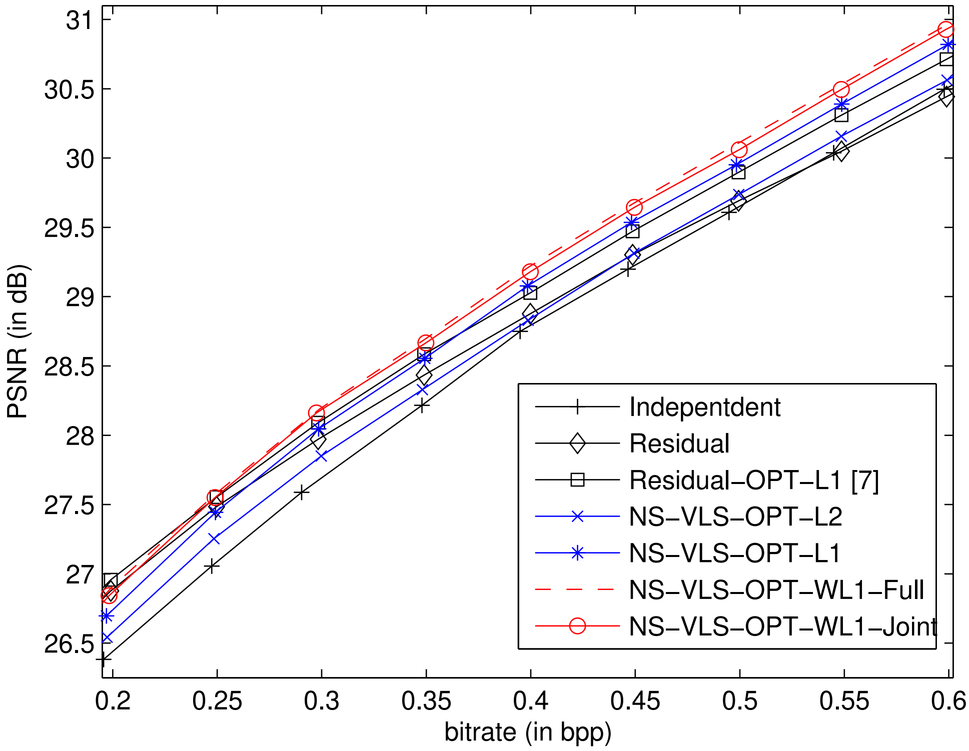


Figure 3.2: *PSNR (in dB) versus the bitrate (bpp) after JPEG2000 encoding for the "houseof" stereo pair.*

Compared to the sparse ℓ_1 optimization, the joint optimization algorithm leads to a gain of about 1-3% and 0.1-0.2 dB in terms of bitrate saving and PSNR, respectively. The gain becomes much more important compared to the standard ℓ_2 optimization algorithm and it reaches 11.6% and 0.85 dB in terms of bitrate saving and PSNR, respectively.

Finally, we have evaluated the proposed joint optimization algorithm in terms of visual quality of reconstruction. Figs. 3.4 and 3.5 display the reconstructed (i.e decoded) target images, for the "art" and "dolls" stereo pairs, using joint optimization algorithm, the standard independent ℓ_2 one as well as the state-of-the-art residual based coding method. Their corresponding PSNR and SSIM [128] values are also provided. Note that a zoom is applied to the reconstructed images to better illustrate the differences between them. It can be noticed that the residual-based coding method may lead to some blocking artifacts. This

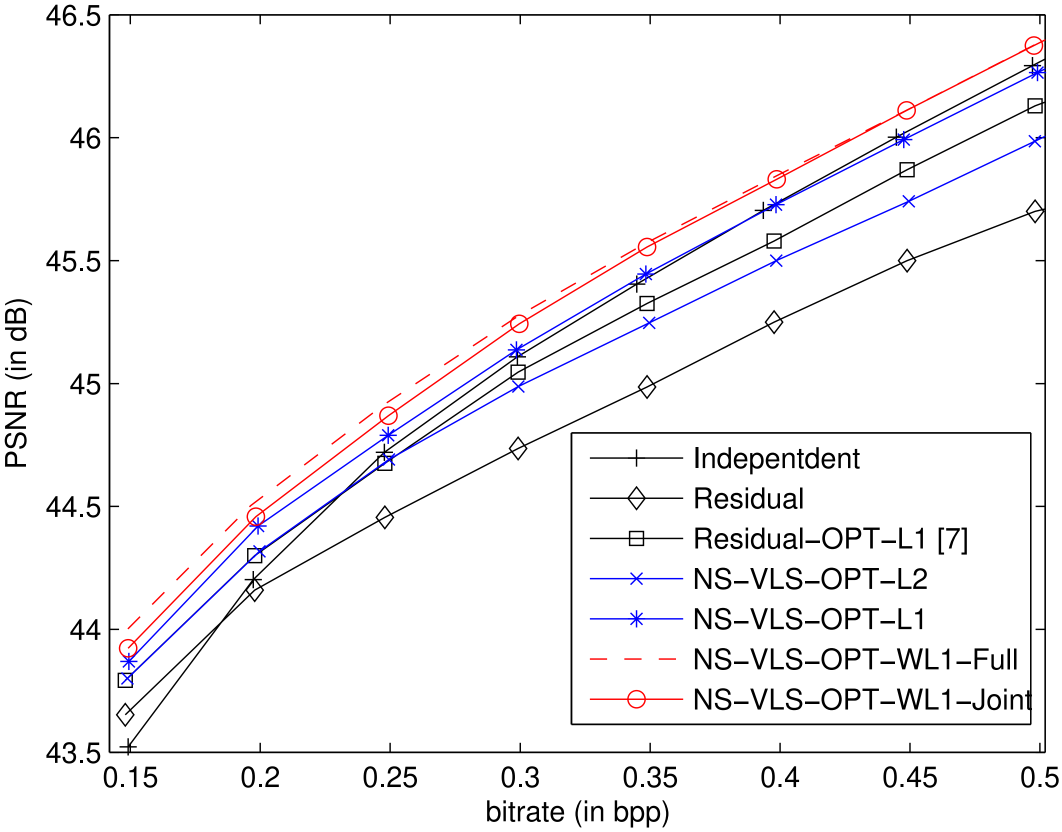


Figure 3.3: *PSNR (in dB) versus the bitrate (bpp) after JPEG2000 encoding for the "ball" stereo pair.*

Table 3.1: *The average PSNR differences and the bitrate saving at low, medium and high bitrates. The gain of "NS-VLS-OPT-WL1-Joint" w.r.t NS-VLS-OPT-L1.*

	bitrate saving (in %)			**PSNR gain (in dB)**		
Images	**low**	**middle**	**high**	**low**	**middle**	**high**
Houseof	-2.95	-2.30	-1.38	0.10	0.12	0.11
Ball	-1.83	-2.81	-1.98	0.05	0.10	0.11
Moebius	-0.93	-2.28	-1.14	0.05	0.14	0.10
Art	-0.92	-0.66	-0.56	0.05	0.06	0.06
Dolls	-2.14	-1.33	-1.03	0.10	0.10	0.11
Playtable	-2.15	-3.05	-1.93	0.10	0.18	0.19
Piano	-1.90	-1.17	-1.24	0.10	0.09	0.15
Teddy	-1.30	-1.11	-0.95	0.07	0.08	0.08
Jadeplant	-2.34	-1.14	-1.08	0.11	0.10	0.11

problem is reduced by resorting to a NS-VLS decomposition, as it can be seen from the results obtained with "NS-VLS-OPT-L2". Moreover, compared to the latter, the joint optimization method improves again the visual quality while

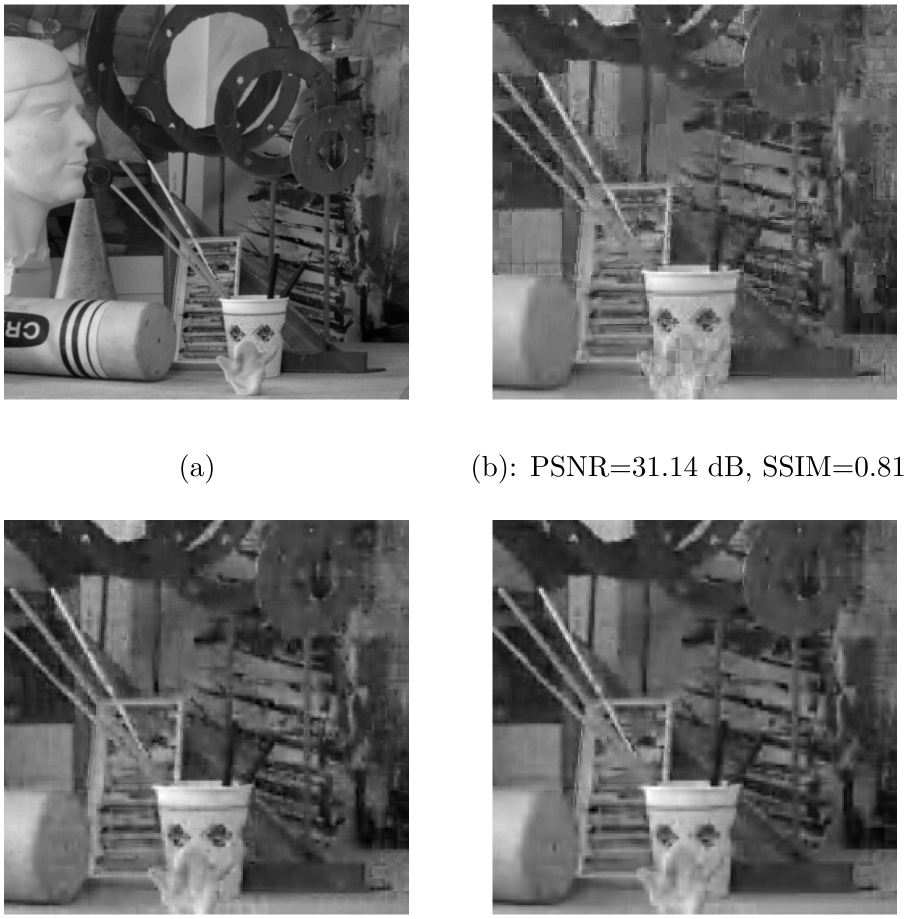
Table 3.2: *The average PSNR differences and the bitrate saving at low, medium and high bitrates*

	bitrate saving (in %)			**PSNR gain (in dB)**		
Images	**low**	**middle**	**high**	**low**	**middle**	**high**
Houseof	-8.50	-7.44	-5.54	0.25	0.38	0.39
Ball	-7.74	-11.67	-6.67	0.14	0.39	0.36
Moebius	-7.35	-8.96	-7.20	0.33	0.55	0.61
Art	-7.78	-9.17	-8.26	0.41	0.64	0.86
Dolls	-7.78	-6.12	-6.45	0.32	0.41	0.67
Playtable	-3.70	-8.29	-7.48	0.17	0.49	0.71
Piano	-7.32	-6.92	-6.13	0.32	0.50	0.71
Teddy	-4.41	-4.81	-4.85	0.18	0.32	0.41
Jadeplant	-6.61	-5.69	-5.06	0.28	0.40	0.50

better preserving the object edges.

3.6 Conclusion

In this chapter, we have investigated different sparse optimization techniques to design the prediction filters of a non separable vector lifting scheme for stereo image coding purpose. In this context, an independent full optimization of the left and right image filters as well as a joint optimization method have been developed. Video coding reduces significantly the amount of data being transmitted over the network. However, in case of wireless multimedia sensor network, traffics still many voluminous and very delay sensitive. In fact, a transmission delay can constitute a major issue when an abnormal event is occurred. Therefore, to avoid this problem, the traffic needs to be scheduled based on its priorities in order to guarantee an efficient video transmission. To this end, in the next chapter, a scheduling model based on priority of traffics for multimedia sensors is proposed.



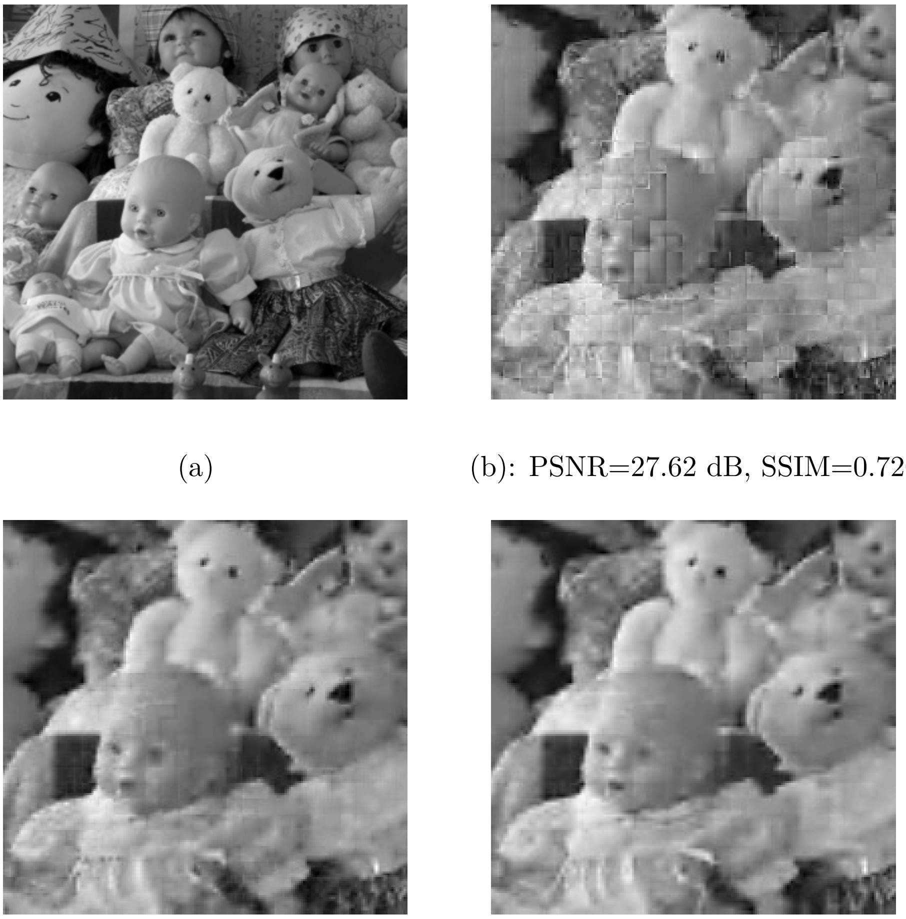
(a)

(b): PSNR=31.14 dB, SSIM=0.818

(c): PSNR=31.04 dB, SSIM=0.829

(d): PSNR=31.45 dB, SSIM=0.848

Figure 3.4: *(a) Original "art" right image. Zoom applied on the reconstructed image at 0.3 bpp using: (b) Residual scheme, (c) NS-VLS-OPT-L2, (d) NS-VLS-OPT2-WL1-Joint.*



(a)

(b): PSNR=27.62 dB, SSIM=0.720

(c): PSNR=28.17 dB, SSIM=0.766

(d): PSNR=28.43 dB, SSIM=0.780

Figure 3.5: *(a) Original "dolls" right image. Zoom applied on the reconstructed image at 0.3 bpp using: (b) Residual scheme, (c) NS-VLS-OPT-L2, (d) NS-VLS-OPT2-WL1-Joint.*

CHAPTER 4

Quality of Service in Wireless Multimedia Sensor Networks

"It's our new method for determining who we should treat first. We take people in order of how loud they scream."

Abstract

The interconnection of all existing communicant devices even the appliances is a new trend of communication that shall empower or renovate the way people interact with each other or with their environment. However, this is a very complicated and demanding task to fulfill the requirements of the transmissions. Some of them are delay sensitive or loss sensitive while others are bandwidth demanding. Considering the case of Wireless Multimedia Sensor Networks (WMSNs) where traffics are many times voluminous and delay sensitive compared to scalar sensors, the problem is twofold: minimizing delay on the one hand and maximizing the throughput on the other hand.

In this chapter, we propose a scheduling architecture model based on priority of traffic consisting of a preemptive and non preemptive priority components and a weighted round robin component. Analytical and simulation results show good performances in terms of minimizing delay for high priority classes and increasing throughput for the overall network.

4.1 Introduction

Nowadays, a new trend of technology is emerging for interconnecting all communicant devices which range from computers, smart phones to home appliances. The Internet of Things (IoT) [129, 130, 131] is that concept that aims to interconnect those devices. Beside this, sensor networks play a great role in devices interconnection. For that purpose and depending on the application domain, many standards in the family of IEEE.802.15 [1] have been released helping accelerate sensor networks development. This communication concerns all traffic types including alarm signals (with a very thin size) and large volume multimedia flows like audio and video data.

For Wireless Multimedia Sensor Networks (WMSNs), the context is particular in the way transmissions and data are managed. In fact, transmissions in WMSNs are managed by taking into account time and loss sensitivity and traffic volume. As we know, multimedia devices especially cameras generate voluminous traffics and in some cases these traffics require a small delay for delivery. The main practical case of this application is the video surveillance systems where some scenarios captured by the cameras require much attention.

Many authors have so far worked on how multimedia traffics are transmitted between small devices while limiting energy consumption and mitigating the waiting time of transmissions especially for data requiring priority in processing. However, the tradeoff between priority and data volume is not enough investigated, what besides comes up in this work.

The concept of scheduling in Wireless Sensor Networks (WSN) has been studied by considering the nature and time sensitivity requirements such as in medical applications [132] or designing scheduling techniques based on real-time requirements [133] or else developing algorithms to increase throughput [134]. In MWSNs, the concept has been also studied by some authors[135, 136, 137, 138, 139] without making an evidence of the influence of multimedia because they

[1] IEEE 802.15 is a standards group of the Institute of Electrical and Electronics Engineers (IEEE) which currently has ten main areas of development

retake the same criteria as in WSN. For example, Alaei et al. [136] and Wang et al. [139] emphasize the scheduling to optimize energy consumption, while Nassim et al. [137] studied the model of WMSNs platform.

Among all these works, no one studied the problem of waiting time for multimedia data requiring low delay. Imane et al. [140] tried to use the priority idea in wireless sensor networks but multimedia data are not involved. Elham K. et al. [141] proposed a priority scheduling mechanism that assigns a correct priority value to a video packet in proportion of other packets in network that have high traffic with no video packet that improves the end-to-end delay and percentage of lost frames. However this is also selective as it deals with video data only. A. Salim et al. [138] proposed a dynamic efficient scheduling strategy for sensor nodes' activities in critical mission of surveillance to optimize network lifetime. Here also, the authors do not consider different priorities of the network flow classes and their requirements in delay and data volume. Still, Shu Fan [142] proposed a joint flow control, routing, scheduling, and power control scheme to increase network lifetime and scheduling fairness but does not really treat the multimedia transmissions especially the different priorities. It is also the case of Nidal N. et al. [143] who proposed a Dynamic Multilevel Priority Packet Scheduling Scheme for Wireless Sensor Network to mitigate end-to-end delay for the highest priority transmissions while keeping acceptable fairness towards the lowest-priority transmissions. This scheme is close to current issue but the problem is still its applicability to the WMSN.

In this chapter, we propose a model of scheduling based on prioritization of multimedia data with delay sensitivity especially in image-video surveillance systems and applications. The contributions of our work are as follows: First, compared to the previous works, ours includes priority in CSMA-CA (Carrier Sense Multiple Access with Collision Avoidance) technique to cope with multimedia (image-video) transmission requirements and categorizes the traffic flows into reasonable classes (this does not prevent from running tests on many more classes). Second, the proposed solution is applicable; it takes into consideration the real problem of WMSN in solution modeling and was designed from a real problem.

Third and finally, rigorous mathematical tools (queue systems) used for analysis is used to optimize throughput and mitigating delay and waiting time.

4.2 Model description

The IEEE 802.15.4 standard [144] gives the details on the packets generated in the network as beacon, data, MAC commands and acknowledgment (ack) packets. In the case of data, we can have other kind of packets such as intermittent, periodic and repetitive traffics. For intermittent traffics, applications define the rate and the device is connected only when it has data to transmit. This has the advantage of optimizing energy as it is the case for Smoke Detectors and Light Switches for instance. For periodic traffics however, the applications set the rate and the transmission occurs following a periodic frequency. It is the case of sensing temperature. Unlike in the aforementioned traffics, the rate is guaranteed and fixed in advance for the repetitive traffics. Medium access for transmission is often managed in slotted time.
All these traffics have different priorities regarding the medium access requirements when we consider the traffic volume, end to end delay or data loss sensitivity.

As we are interested in multimedia communication where traffics have big volume with different priorities, we first propose a new algorithm for accessing the medium; it is a modification of the original CSMA-CA algorithm to cope with priorities. Just next to the proposed medium access algorithm, we set a scheduling architecture comprising a Weighted Round Robin (WRR) scheduler, a Non Preemptive Priority (NPP) scheduler and a Preemptive Scheduler (PS). This scheduling architecture aims to cope with the volume of transmissions reinforced by priorities.

4.2.1 Priority based CSMA-CA

The IEEE 802.15.4 standard allows two kinds of network configuration modes, beacon and non-beacon enabled modes. In beacon-enabled mode, a network coordinator periodically generates beacon frames after every Beacon Interval

(BI) in order to identify its network to synchronize with associated nodes and to describe the superframe structure. This superframe is only used in the beacon-enabled mode. Regarding the non-beacon-enabled mode, all nodes can send their data by using an unslotted CSMA/CA mechanism. This mechanism does not provide any time guarantees to deliver data frames.

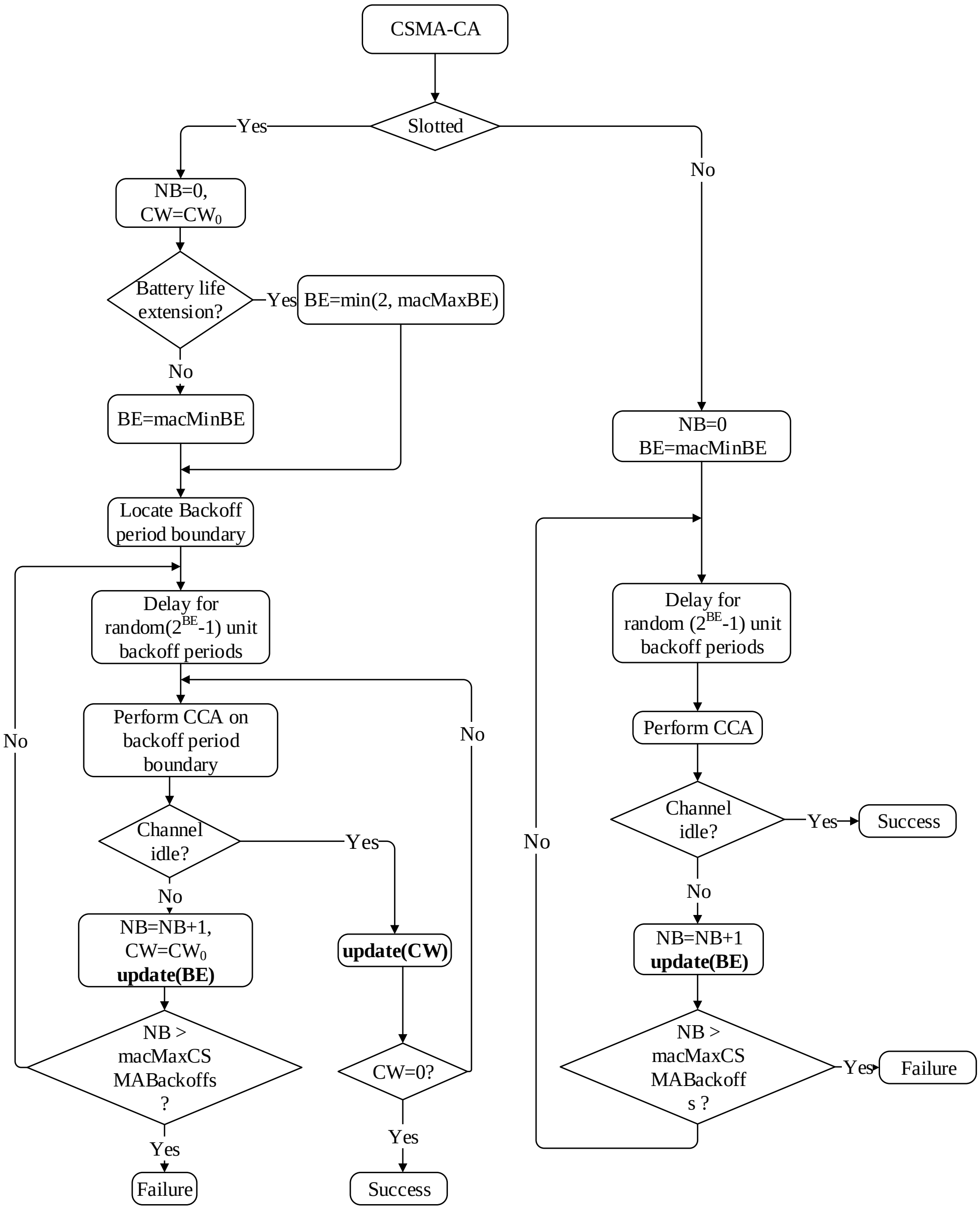


Figure 4.1: *Priority CSMA-CA algorithm*

The problem in the original CSMA-CA stands at the fact that all traffics are processed in the same way regarding their medium access attempts for data transmission. However, when priorities are set without any change the traffics with high priority are penalized. Thus, in this work we have provided changes in the way backoff exponent (BE) and contention window (CW) values are updated and changed to integrate the priority effect. These changes follow Eq. 4.1 for **update(BE)** and Eq. 4.2 for **update(CW)**.

$$BE_k(P_i) = min(BE_k(P_i) + \frac{k + BE_k(P_i)}{P_i}, macMaxBE) \tag{4.1}$$

$$CW_k(P_i) = \begin{cases} 0, \text{when } CW_k(P_i) < \frac{2k+P_i}{CW_k(P_i)} \\ CW_k(P_i) - \frac{k+P_i}{CW_k(P_i)}, \text{when } k \quad \text{is odd,} \\ CW_k(P_i) - \frac{2k+P_i}{CW_k(P_i)}, \text{when } k \quad \text{is even} \end{cases} \tag{4.2}$$

where

- k stands for the number of transmission retries a node has already done for the transmission of a data packet

- P_i is the priority of the traffic class i and

- $macMaxBE$ is the maximum value of the backoff exponent which varies from 3 to 8 and its default value is 5 [144].

These updates have the advantages of accessing quasi immediately the medium for traffics with high priority. For example, the BE value will increase slowly so as to allow many attempts to transmit and the CW value will decrease so fast and reaches zero what will make the immediate transmission. However, for traffics with low priority the medium access will stay normal.These updates are integrated in the CSMA-CA general diagram [144] and the modified diagram is given in Figure 4.1.

The performances of the modified CSMA-CA algorithm according to the Eq. 4.1 and Eq. 4.2 are illustrated in Figures 4.2 and 4.3. As depicted in Figure 4.2, the waiting time before sending increases following the decrease of the priority value, that is, the traffic flows with high priority have small waiting time, contrary to the standard CSMA-CA algorithm where all traffic flows have the same waiting

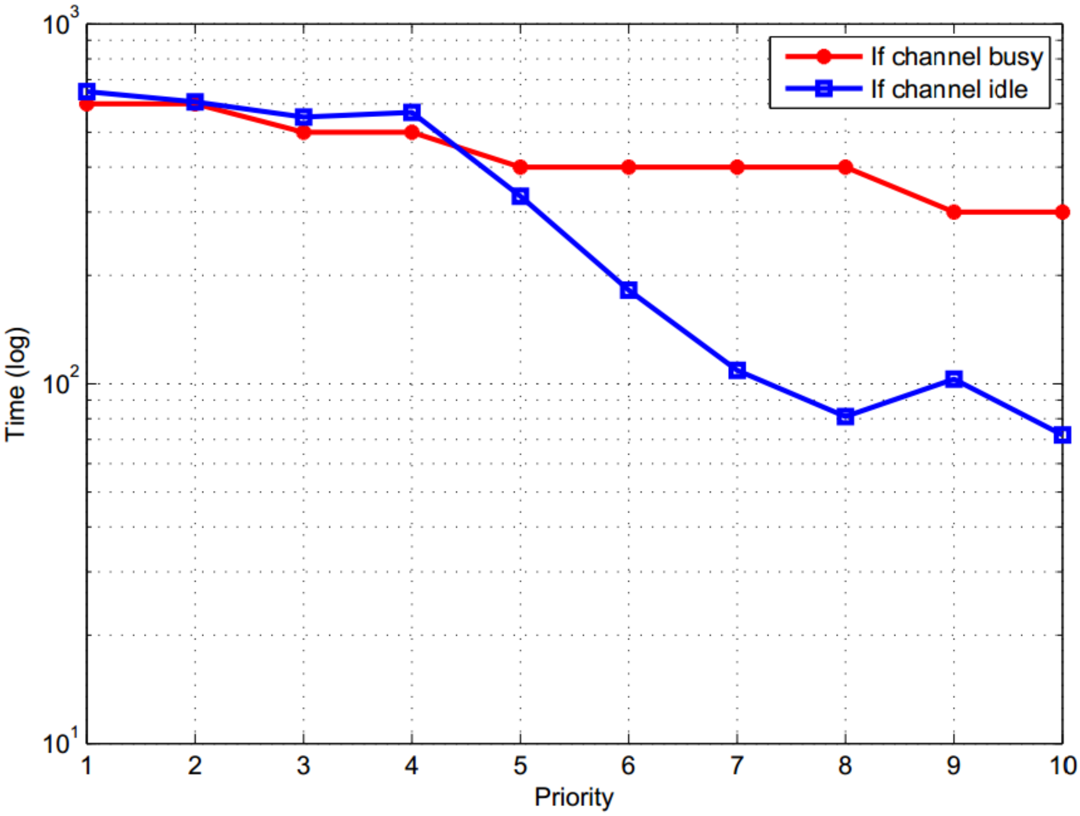


Figure 4.2: *Waiting time*

time. Regarding the number of transmission tries, Figure 4.3 shows that traffics of high priority make many attempts trying to access the medium when the channel is idle (before NB, the number of backoffs or periods, value reaches maximum value, $macMaxCSMABackoffs$ which is the maximum NB value the CSMA-CA algorithm will attempt before declaring a channel access failure [144]) and very few attempts when the channel is busy (the contention window value decreases rapidly).

4.2.2 Traffic classes

In WMSNs, traffics are especially image and video and except that they are a high bandwidth consumers, they are different from each other in terms of priority due to some points of view. On the one hand, video source is the first parameter for setting priority. For example, video from surveillance cameras in a nuclear site or a weapons store drain much attention than those from a public place. Thus, the former has higher priority than the latter. In this way, since several cameras

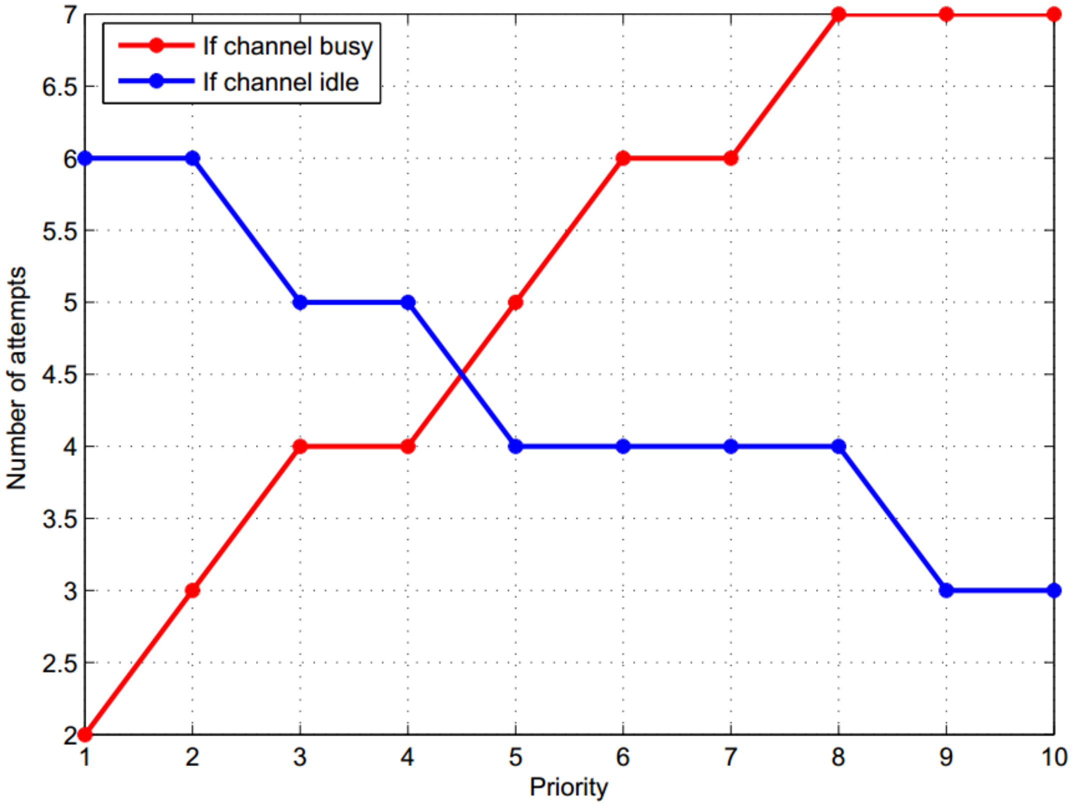


Figure 4.3: *Tries*

can be set in a same place they will have different priorities given their sources.

One can define the priority levels according to the number of sources or the source sensitivity as a given number of sources can have the same data sensitivity. Let n be the number of places where cameras are installed and N the whole amount of cameras, respectively. The number of priorities $Card(P)$ to be set satisfies the relation $Card(P) \leq n \leq N$.

On the other hand, the way video traffic is generated is also another parameter of classification and then grants priority. There can be many more classes but three are reasonably possible.

Periodic or Repetitive Flow (PRF): this traffic is generated at regular time slots defined in advance. The source is configured to send traffics t periods of time; it is a *store-and-forward* mechanism. For example, a camera takes video without sending and starts sending when the period of time t ends up.

On-Demand Flow (ODF): regarding this traffic, the video flow is transmitted only on demand. It is the case when a place is suspected and the video from that place's cameras is asked. Assume that a camera is set to give the previous traffic. Even before the sending time comes up, a video traffic can occur under sending demand.

Event Driven Flow (EDF): this traffic is intermittent and related to the one generated when there is something important. For example when a surveillance camera is in idle state and wakes up to take video because something passes in its sight.

These three traffic classes have different priorities. Let $P(X)$ be the priority of the traffic X. The priorities of the traffics PRF, ODF and EDF satisfy the following inequality: $P(PRF) < P(ODF) < P(EDF)$

4.2.3 Use case

The previous model has suitable practical use case for example by considering the Institut Galilée campus at the university paris 13. There are surveillance cameras at the car parking, at all the entries and at library. All these cameras generate traffic flows with different priorities. For instance, flows from car parking cameras have higher priority than those from different entries.

Considering the settings from the Figure 4.4, PRF traffics are generated by cameras located at A and B, EDF traffics are generated by cameras located at B and C while ODF traffics are from cameras at A, B and C. These settings will be used later as test-bed for simulations.

4.3 Scheduling

There are many scheduling strategies such as Weighted fair queuing (WFQ), Low Latency Queuing (LLQ),Round Robin (RR) [145], weighted RR [132, 146] etc. some of them have been studied [147, 148].However, William Stallings in his book [149] made an interesting comparative study where RR seems to meet many criteria. In this way, the idea of using it in our model came up. Thus, by taking the priorities of classes as weights leads to WRR strategy.

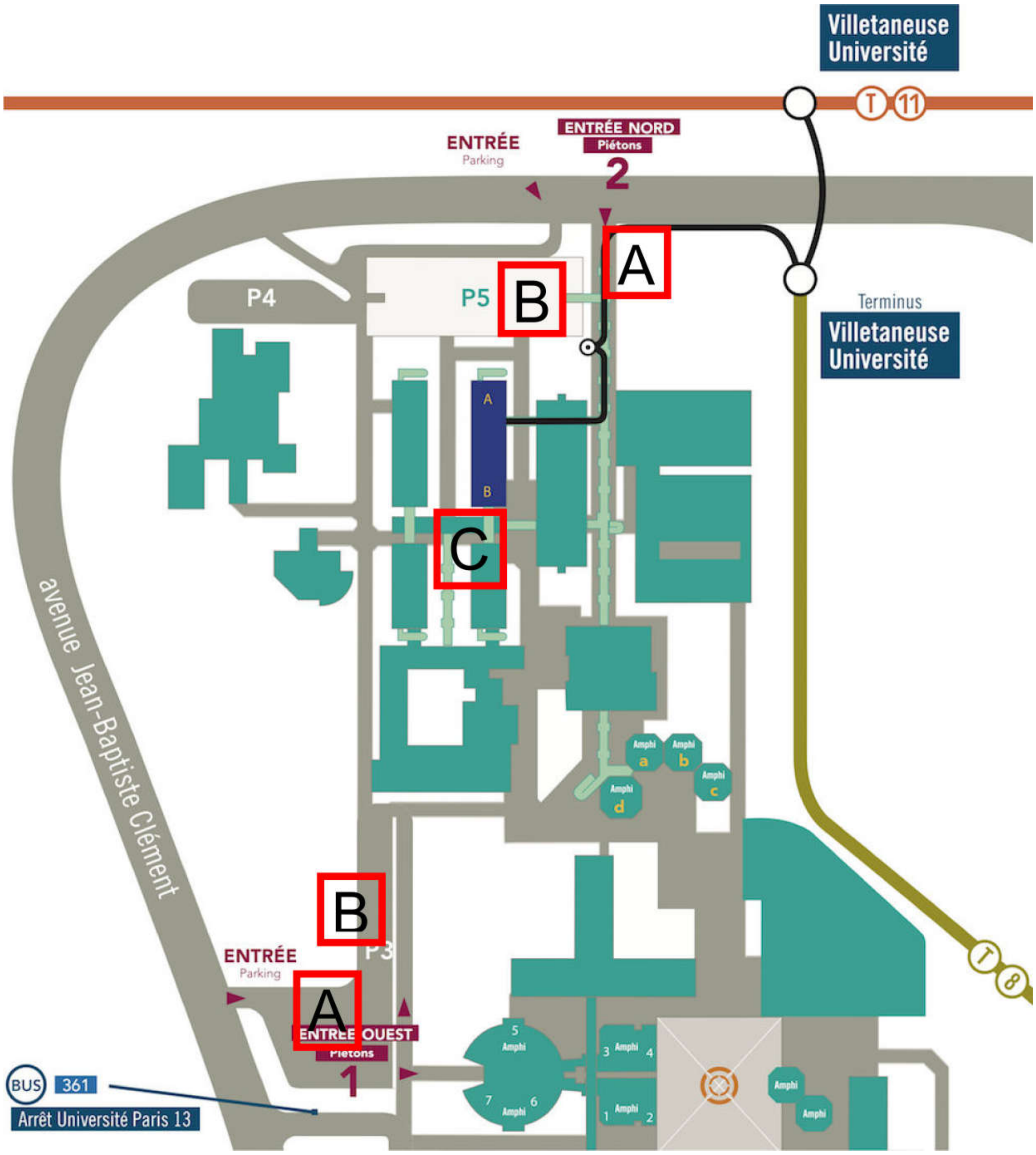


Figure 4.4: *View of Institut Galilée - Paris13 University*

As depicted by Figure 4.5, two stages of scheduling are necessary due to different priorities that can have the traffics given their types and sources. The first stage is related to EDF and the scheduler policy is weighted round robin, WRR [150, 151]. At this stage there is another scheduler, a non preemptive priority (NPP) scheduler for ODF and PRF according to their priorities. The second stage concerns all the outputs of WRR and NPP using a preemptive priority (PP) scheduler for a general output.

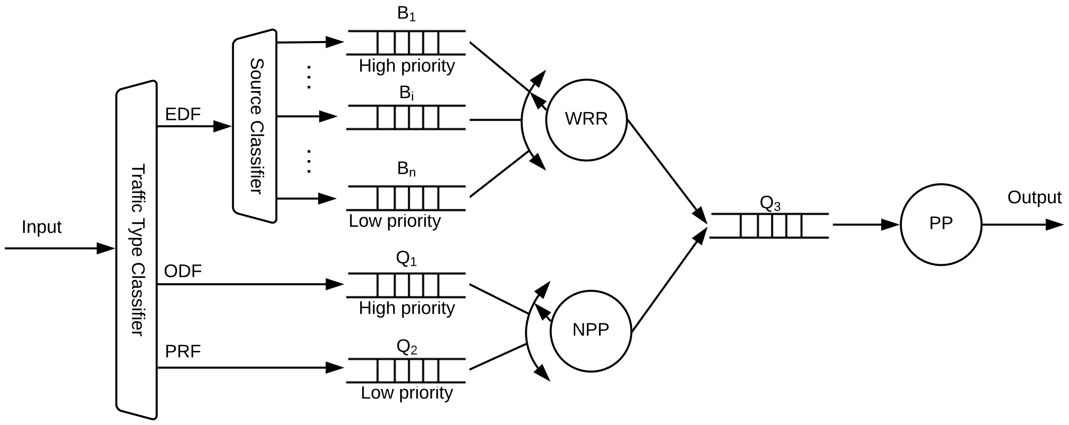


Figure 4.5: *Scheduling architecture for WMSN*

Let's consider the M/G/1 queuing model[152] with n customer classes having n priorities (1 is the highest, n is the lowest), the arrival rate of the i^{th} class follows the Poisson process with rate λ_i and service times $S_i = \frac{1}{\mu_i}$ and $S_i^2 = \frac{1}{\mu_i^2}$ generally distributed. We also consider a steady and stable state with

$\rho_i = \frac{\lambda_i}{\mu_i}$

$\rho = \rho_1 + \rho_2 + ... + \rho_n$ with $\rho < 1$

where ρ_i denotes the fraction of time allocated to the class i by the server and ρ represents the time the server is busy and thus the time the server available or idle is given by $1 - \rho$. As for μ, it represents the total service rate of all classes with $\mu = \mu_1 + ... + \mu_n$.

Let $E(W_i)$ be the mean customers waiting time of class i, $E(L_i)$ the mean waiting customers number of class i and $E(R)$ the mean residual service time.

4.3.1 Non Preemptive Priority scheduling, NPP

For G/M/1 NPP model, the scheduling is only priority driven service and the queue is like the single queue with G/M/1 FIFO (First In First Out) policy with priority because the service of each flow waits for all higher priorities to complete their service unless it is under service.

Considering that non-preemptive regime, let's compute the waiting time of the classes.

Given that the priority i is higher than $i+1$, we get by the Pollaczek Khinchine Formula for the high priority class [153, 154]:

$$E(W_1) = E(R) + \frac{1}{\mu_1} E(L_1) \tag{4.3}$$

Applying the Little's law ($E(L_i) = \lambda_i E(W_i)$[153][155]), we get:

$$E(W_1) = \frac{E(R)}{(1-\rho_1)} \tag{4.4}$$

For the class of second priority, we get:

$$E(W_2) = E(R) + \frac{1}{\mu_1} E(L_1) + \frac{1}{\mu_2} E(L_2) + \frac{\lambda_1}{\mu_1} E(W_2) \tag{4.5}$$

With Little's Law, we always get:

$$\begin{aligned} E(W_2) &= E(R) + \rho_1 E(W_1) + \rho_2 E(W_2) + \rho_1 E(W_2) \\ &= (E(R) + \rho_1 E(W_1))/(1-\rho_1-\rho_2) \end{aligned} \tag{4.6}$$

By applying Eq. 4.4in 4.6, we get:

$$E(W_2) = \frac{E(R)}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{4.7}$$

By generalization, let's calculate the $E(W_i)$ with $i > 0$

$$\begin{aligned} E(W_i) &= E(R) + \frac{1}{\mu_1} E(L_1) + ... + \frac{1}{\mu_i} E(L_i) + \frac{\lambda_1}{\mu_1} E(W_2) + ... + \frac{\lambda_{i-1}}{\mu_{i-1}} E(W_i) \\ &= E(R) + \rho_1 E(W_1) + ... + \rho_{i-1} E(W_{i-1}) + (\rho_1 + ... + \rho_i) E(W_i) \end{aligned} \tag{4.8}$$

By induction, we finally get:

$$E(W_i) = \frac{E(R)}{(1-\rho_1-...-\rho_{i-1})(1-\rho_1-...-\rho_i)} \tag{4.9}$$

Let ρ_i^+ be the sum of the series. Then the Eq. 4.9 becomes as Eq. 4.10

$$E(W_i) = \frac{E(R)}{(1-\rho_{i-1}^+)(1-\rho_i^+)} \tag{4.10}$$

The residual waiting time $E(R)$ [153] is given by

$$E(R) = \frac{1}{2}\lambda E(S^2) \tag{4.11}$$

Due to the following expressions: $E(S^2) = 1/\mu^2$ and $\rho = \lambda/\mu$

$$E(R) = \frac{\rho}{2\mu} \tag{4.12}$$

Therefore, the final expression of waiting time is given by Eq. 4.13 after incorporating (4.12) in (4.10) :

$$E(W_i) = \frac{\rho}{2\mu(1-\rho_{i-1}^+)(1-\rho_i^+)} \tag{4.13}$$

4.3.2 Preemptive Priority scheduling, PP

For PP scheduling, the model is like in previous NPP except that the traffic flows are penalized with the arrival of traffic flows of high priority.
Let's calculate the waiting time in this case: we know that the class i does not see classes $i+1,...,n$; i.e the classes of low priority than itself.

Recall for M/G/1-like queues [153], $E(W) = \frac{E(R)}{1-\rho}$ generalized by Eq. 4.9.

$$E(W_i) = \frac{E(R)}{(1-\rho_i^+)} + \sum_{j=1}^{i-1} \lambda_j \frac{1}{\mu_j}(E(W_i) + \frac{1}{\mu_i}) \tag{4.14}$$

$\frac{1}{\mu_i}$ is because all traffic flow $j \leq i-1$ preempts i.

$$\begin{aligned} E(W_i) &= \frac{E(R)}{(1-\rho_i^+)} + \rho_{i-1}^+ (E(W_i) + \frac{1}{\mu_i}) \\ &= \frac{E(R)}{(1-\rho_i^+)(1-\rho_{i-1}^+)} + \frac{\rho_{i-1}^+}{(1-\rho_{i-1}^+)} \frac{1}{\mu_i} \end{aligned} \tag{4.15}$$

As the i^{th} flow will be only interrupted by all $j \leq i-1$ flows, the residual waiting time $E(R)$ is defined by the sum of residual waiting times of classes from the highest up to the current class.

Thus the $E(R)$ is given by the expression

$$E(R_i) = \frac{1}{2} \sum_{j=1}^{i} \frac{\lambda_j}{\mu_j^2} = \frac{\rho_i^+}{2\mu_i^+} \tag{4.16}$$

Therefore the general expression is given by Eq. 4.17

$$E(W_i) = \frac{\rho_i^+}{2\mu_i^+(1-\rho_i^+)(1-\rho_{i-1}^+)} + \frac{\rho_{i-1}^+}{(1-\rho_{i-1}^+)} \frac{1}{\mu_i} \tag{4.17}$$

4.3.3 Weighted Round Robin, WRR

The Round Robin (RR) is a scheduling technique that cyclically selects an equal amount of data (packets) from each queue buffer for transmission. This way of doing ensures equality of traffics but is not without drawbacks. For example, traffics with big volume or time sensitivity will be penalized. Hence, the WRR is a solution. The WRR scheduling is a modified RR scheduling that adds weights to different queues present in the system and the WRR scheduler takes amount of packets as per the weight assigned to each queue. In this way, it prevents packets of lower priority queues from starvation while struggling to meet the requirements of minimizing delay. In the case of this work, the WRR uses the priority for weights.

The WRR scheduling is based on assigning fraction weight φ_i to each service queue such that the sum of weights of all service queues is equal to one.

$$\sum_{i=1}^{n} \varphi_i = 1 \tag{4.18}$$

Considering the M/G/1 model, the WRR is like NPP except that in the former the flow from a given class must wait until the scheduler rounds on all queues. In this way, the waiting time increases according to the priority and the probability of finding waiting flows in the queues.

From then, the waiting time of a traffic flow of a given class is the product of the possible number of rounds in the current class with the sum of all waiting times of other classes.This is because in WRR, the traffic flow of each class waits until the scheduler serves all other classes and the traffics flows of the same class. The number of rounds $\Re$ is calculated by considering the expected number of customers in the current class and its service time. Thus, this number is given by :

$$\Re = \frac{E(L_i)}{\mu_i} \tag{4.19}$$

By Little's law, the Eq. 4.19 becomes as follows:

$$\Re = \rho_i E(W_i) \tag{4.20}$$

From then on, the waiting time is the residual time plus the product of traffic intensity of each class with the number of rounds, what gives

$$\begin{aligned} E(W_i) &= E(R) + \Re(\rho_1 + \rho_2 + ... + \rho_n) \\ &= E(R) + \Re\rho \end{aligned} \tag{4.21}$$

By replacing Eq. 4.20 in Eq. 4.21, we get:

$$\begin{aligned} E(W_i) &= E(R) + \rho_i E(W_i)\rho \\ &= \frac{E(R)}{1-\rho_i\rho} \end{aligned} \tag{4.22}$$

The residual waiting time is the remaining service time of the job already in service. In WRR, however, this time depends on which class the traffic flow in service is from and the number of classes. In this way, this waiting time is given by

$$\begin{aligned} E(R) &= \frac{1}{2}n\lambda_i E(S_i^2) \\ &= \frac{n\rho_i}{2\mu_i} \end{aligned} \tag{4.23}$$

Combining Eq. (4.23) and (4.22) gives the final expression (4.24) of the i^{th} traffic flow's waiting time.

$$E(W_i) = \frac{n\rho_i}{2\mu_i(1-\rho_i\rho)} \tag{4.24}$$

4.3.4 First In First Out, FIFO

The FIFO policy is synonymous of absence of scheduling policy and the IEEE 802.15.4 uses the same policy. Then, the waiting time is given by Eq. 4.25 [153] given that all traffic flows are in the same queue without any priority.

$$\begin{aligned} E(W) &= \frac{\lambda E(S^2)}{2(1-\rho)} \\ &= \frac{\rho}{2\mu(1-\rho)} \end{aligned} \tag{4.25}$$

This will serve as comparison reference with the three proposed models.

4.4 Analytical and simulation results

The proposed scheduling architecture is likely to mitigate the waiting time especially for data with high priority and thus maximize the throughput. The following are the settings for analytical computation: we set the same value of λ for all classes, and the value of μ increases according to the priority. The index with 1 has the greatest priority where 10 has the lowest while keeping the inequality $\rho < 1$ true.

From the results performed in Matlab and depicted in Figure 4.6, the PP policy is good for only the highest priority class but a worst choice for low priority classes. In fact, for the highest priority class, the waiting time is almost null but the increase becomes quasi exponential. The NPP, however, has good performances compared to PP as for the first classes the waiting time increases slightly; this is due to non preemptive behavior. The WRR comes as the tradeoff in keeping the waiting time variation small for all classes.

The combination of all these three scheduling policy in the proposed model architecture is justified by the need of giving an almost zero waiting time to the delay sensitive traffics. Besides, Figure 4.7 shows that when the number of classes decreases, the waiting time decreases too; what fits to the three classes found in WMSNs.

Figure 4.8 illustrates the mean waiting time in all components of the proposed model. Compared to FIFO model of the standard, the proposed model has good performances if we have a system not exceeding 6 priority classes.

By minimizing the waiting time of traffics, it is obvious that the throughput increases in the same period of time as illustrated by Figure 4.9. This is also confirmed by simulation results depicted in Figure 4.11.

For simulations in NS3 [156][157], 5 nodes have been used: one D for destination, another one R for relay and the left 3 S1, S2 and S3 are sources (see Figure 4.10). We considered only three traffic classes according to the traffic classification made in Section 4.2.2. The source nodes S1, S2 and S3 generate packets of EDF, ODF

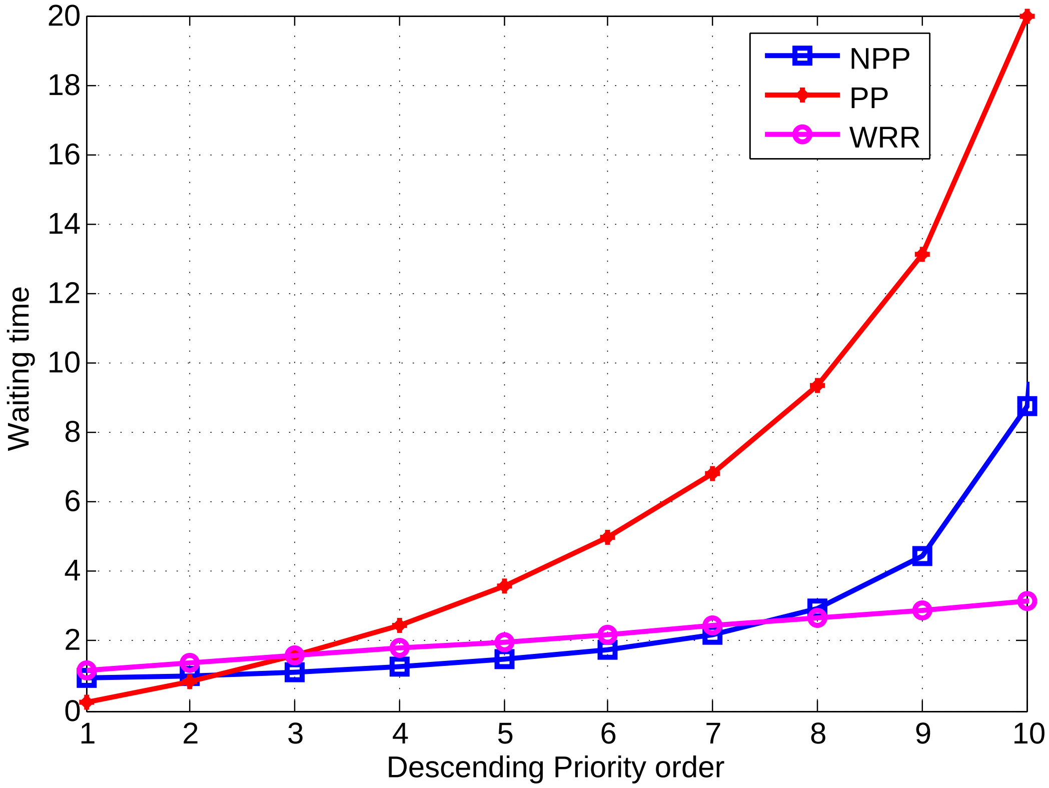


Figure 4.6: *Waiting time comparison*

and PRF classes with respectively 7, 5 and 2 as priority values. They use 802.15.4 model for transmissions and each node sends a packet of 512 octets with a same data rate of 30 transmissions per second.

The results after one minute of simulation running are presented in Figures 4.11 and 4.12. It is clear that even the simulations have the same increase profile of waiting time and the same decrease profile of throughput following the priority order while keeping the same performances over the standard model.

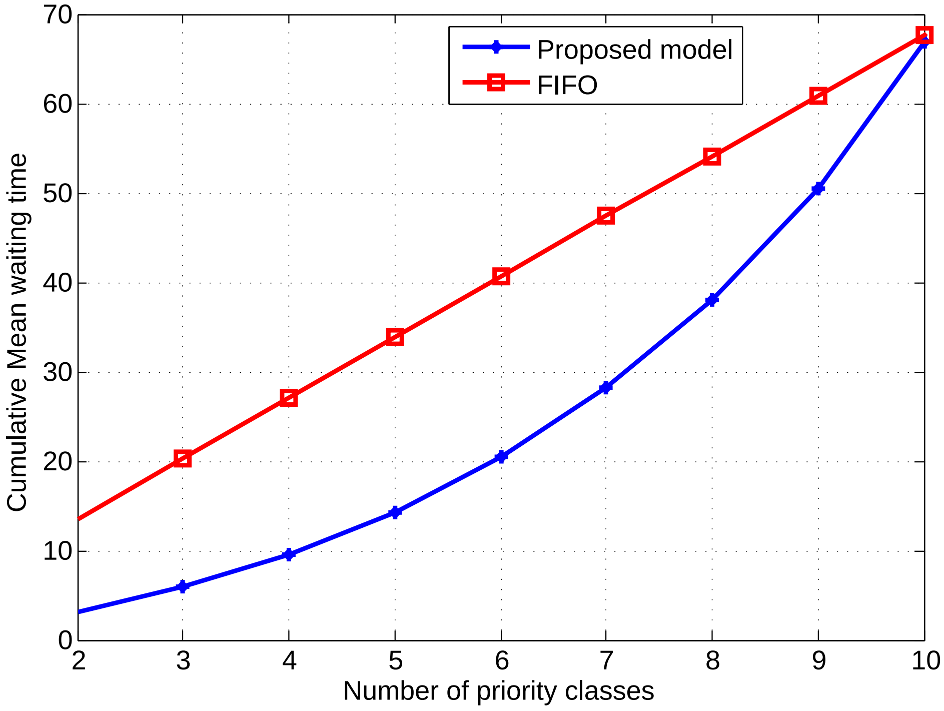


Figure 4.7: *Cumulative Mean waiting time*

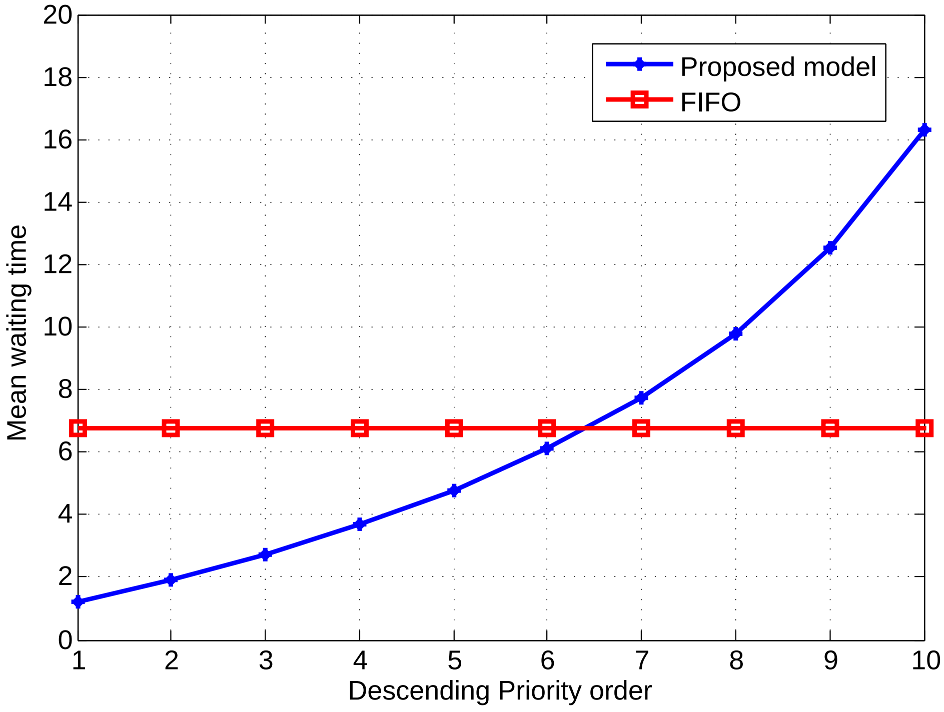


Figure 4.8: *Mean waiting time*

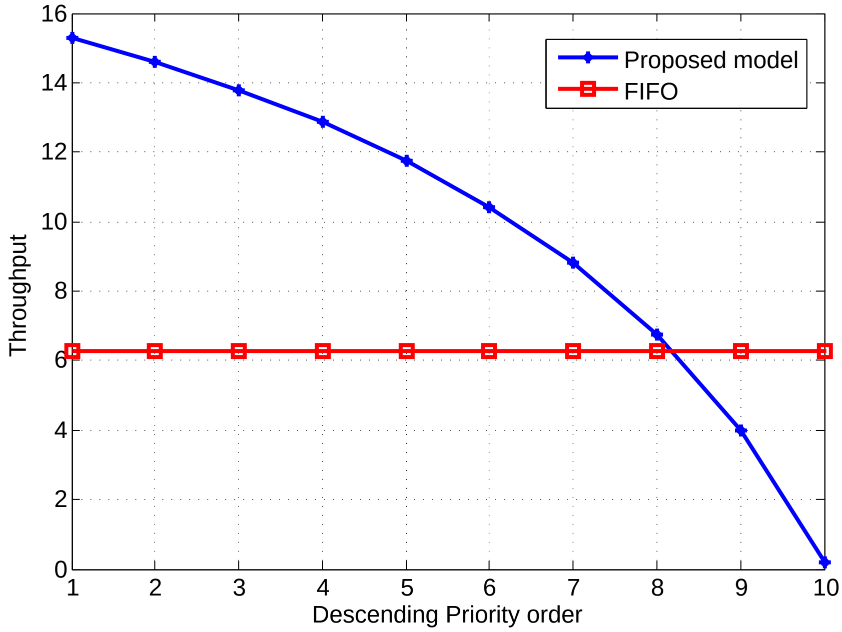


Figure 4.9: *Throughput profile following priority decrease*

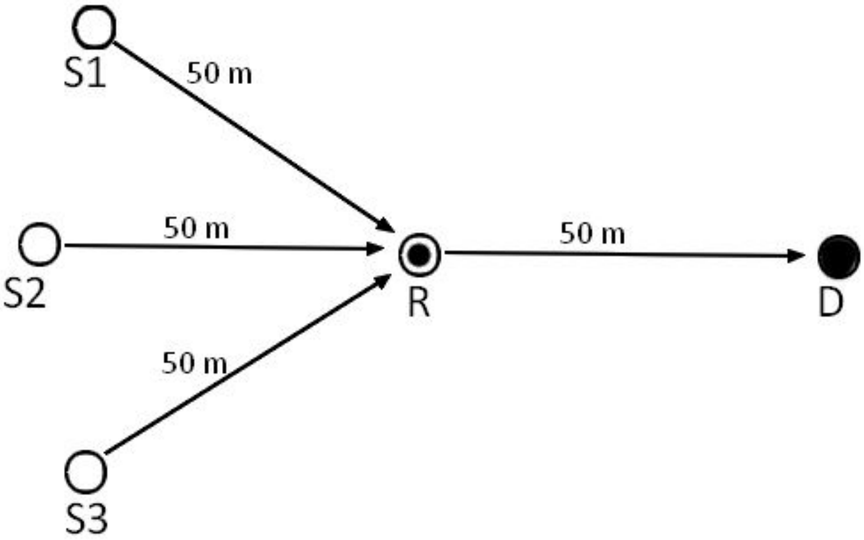


Figure 4.10: *Simulation settings in NS3*

4.5 Conclusions

In this chapter, a scheduling model based on priority of traffics for multimedia sensors is proposed. First, the CSMA-CA model has been modified so as to integrate priority in medium accessing, then the traffic has been classified to comply with the transmission environment as real as possible and finally the proposed scheduling model was set as a combination of weighted round robin and

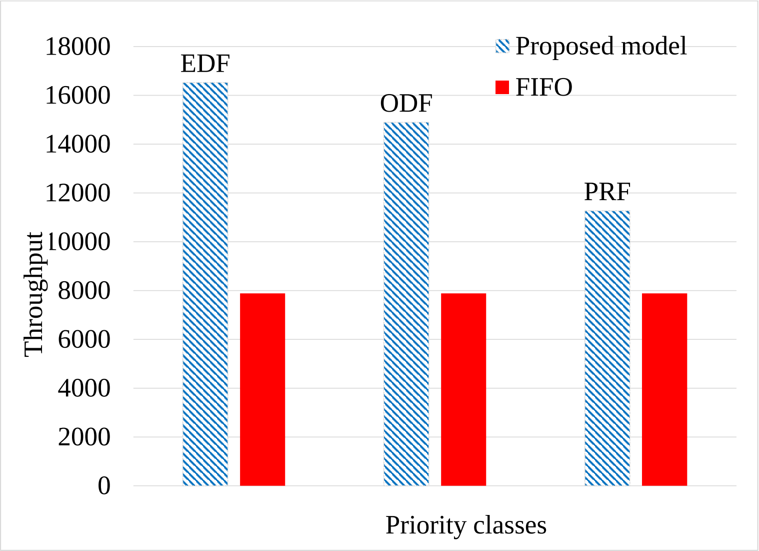


Figure 4.11: *Throughput of EDF, ODF and PRF classes*

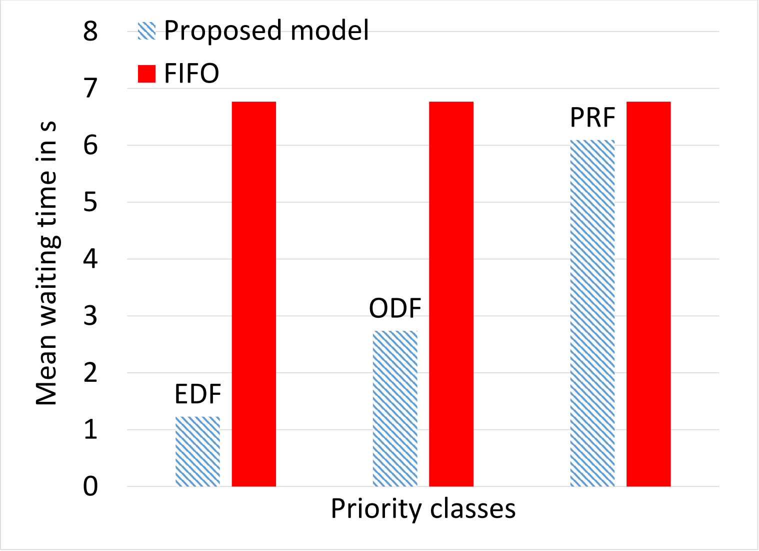
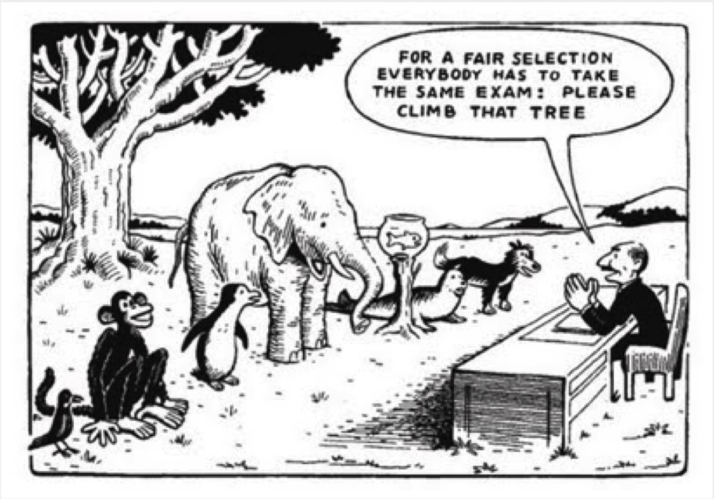


Figure 4.12: *Mean waiting time of EDF, ODF and PRF classes*

strict priority scheduling. The analytical and simulation results show that the proposed model is efficient in mitigating the waiting time while optimizing the throughput following the priority profile if compared with the existing model. This scheduling model guarantee an efficient transmission of the video based on its priority. However, transmission errors can occur and can affect the quality of the transmitted video. In addition, the global video surveillance system's quality will be decreased when a network congestion is occurred. In addition, video coding can affect the video quality which may cause an inefficient object detection and thereafter bad abnormal event detection. Network quality and the video coding process within the captured environment that may cause video distortion are the three factors that affect the received video. Therefore, in order to have an efficient video surveillance system, a video quality assessment task must be performed to evaluate the video quality and to decide whatever or not it needs an enhancement. In the next chapter, we will propose a quality-based intelligent video surveillance system and a video surveillance oriented video quality database.

CHAPTER 5

VQA and Intelligent video surveillance system


FOR A FAIR SELECTION EVERYBODY HAS TO TAKE THE SAME EXAM: PLEASE CLIMB THAT TREE

"Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid."

-Albert Einstein

Abstract

The interconnection of all existing communicant devices even the Video based security surveillance systems often produce poor-quality video in uncontrolled environment. This may strongly affect the performance of high level tasks such as scene understanding and interpretation. The aim of this work is to show the impact and the importance of video quality in smart video surveillance systems. In this chapter we present a video surveillance video quality database that we have created in order to help the scientific community working on video quality assessment and especially for video surveillance application. Then we mainly propose a complete quality-based smart surveillance system and we focus on the most important challenges and difficulties.

5.1 Introduction

In the past decade, the world has witnessed a massive deployment of computational intelligence in video surveillance systems thanks to an intensive research in intelligent analysis of visual information [158] [159]. the old video surveillance systems with human resources dedicated to observe the cameras 24 hours daily has shown their limitations in terms of cost and human failures. Some monitoring activities like abnormal event detection, face detection/ recognition and identification can be done without human intervention thanks to the smart video surveillance systems. Thus, scientists have intensely contributed on the development of some sophisticated algorithms able to detect and recognize the faces under some known conditions related to the video quality.
Video quality is a primary factor on video surveillance process. Indeed, under some weather or material conditions, some video surveillance functions and features like object/face detection and recognition would be very tough tasks. Therefor, the use of some metrics to evaluate the video quality before the processing phase is highly recommended to reach high efficiency.

Quality Assessment research strongly depends upon subjective experiments to provide calibration data as well as a testing mechanism. After all, the goal of all QA research is to make quality predictions that are in agreement with subjective opinion of human observers. In order to calibrate QA algorithms and test their performance, a data set of videos whose quality has been ranked by human subjects is required.
Many algorithms have been developed by the research community, in order to compute objective metrics which try to automatically and reliably predict the quality of the multimedia content as perceived by a human end-user. Objective quality metrics available in literature can be divided in three different categories according to the availability of the original video which are Full-Reference (FR) ,Reduced-Reference (RR) metrics and No-Reference (NR) metrics. These categories will be described in the next section. It is important to note that when a

real-time in-service quality evaluation is needed, only NR metrics can be applied because the unimpaired video is unavailable at the receiver end.

Although human beings can rather easily make judgments about the quality of multimedia contents, predict the subjective assessment by mean of objective algorithms represents a much harder task. The prediction accuracy of the objective metrics is measured by comparing the values of the objective measures with the results of subjective experiments. In fact, subjective tests are usually performed, where a group of individuals is asked to rate the quality of the digital material or the overall multimedia experience. These results are used as benchmark for the objective metrics. However for video context, few are the data-sets that were created for quality assessment. Indeed, researchers use image quality assessment database to test their video quality assessment metrics which does not provide good results.

Many databases have been created for error transmission and noise impact on video quality[160, 161, 162, 163, 164, 165], these databases are very useful for researchers working on video quality impacted by transmission errors. Authors in [162, 163] created a publicly available database of 156 video streams, encoded with H.264/AVC and corrupted by simulating the packet loss due to transmission over an error-prone network.

In literature no video database was created to deal with distortions that impact video in visual surveillance context. Thus a video surveillance data set is primary for our project to test and train the upcoming video quality assessment and enhancement algorithms for visual surveillance.

Recent sophisticated video surveillance systems guarantee good and high object/face or event detection, recognition and tracking. Therefore, a good video quality is needed to ensure this process efficiently. However, in such system different signal processing stages can affect the acquired video quality; capture, network transmission, video compression, causing video distortions. Moreover, video surveillance systems are often deployed in outdoor structures; stadium, buildings, parking and streets but to name few. Based on the outdoor environment nature, some natural degradation due to adverse weather conditions such haze, fog and smoke in some cases greatly reduce the visual quality of outdoor surveillance videos and make the event/object detection and recognition process not efficient. Therefore, any video quality degradation impacts directly the efficiency of the video surveillance system which may straightly affect security. As it impacts directly the efficiency of the object/event detection recognition process, quality has been always considered as the major concern for video surveillance systems. Therefore, video quality assessment (VQA) metrics and methods are primary for video surveillance systems in order to evaluate the video quality before proceeding to the visual detection and tracking process.

Smart video surveillance systems rely on video analysis algorithms for automated video processing. However, little is known about the minimum video quality required to ensure an efficient performance of these algorithms. In [166], authors have been concluded that the algorithms show almost no decrease in accuracy and efficiency until a certain critical quality of the input video is reached, which amounts to significantly lower bit-rate compared to the quality commonly acceptable for human vision. Authors in [167], have shown the effect of bad/uneven illumination on the face detection process. This illumination issue may degrade the perceptual video quality and affect the efficiency of the surveillance process.

5.2 Video Quality Assessment

5.2.1 Subjective quality assessment

Subjective quality assessment represents all psycho-visual experiments in which a number of observers evaluate a given group of stimuli. At the end, a quality score is assigned to each image or video. These tests need a long time to be performed and require a big number of expert and non expert observers. However, they remain the most reliable method of quality assessment since the human observer is always the final receiver of the visual content.

5.2.1.1 Subjective quality protocols

Since the human being is the ultimate judge of perceptual quality. Subjective tests remain a reliable and indispensable tool objective metrics evaluation. In order to build a video database that can be used to evaluate the performance of the proposed metrics certain rules must be followed. The Video Quality Experts Group (VQEG), a group formed by experts from the International communication union ITU, has put forward a number of recommendations to better assess the quality of images/videos subjectively .

Absolute Category Rating (ACR)

The Absolute Category Rating (ACR) method, also known as the single-stimulus method, is a category judgment where test sequences are presented one by one in order to be independently rated on a known category scale. The method specifies that after each presentation, observers are asked to rate the quality of the shown sequence.

Absolute category rating with hidden reference (ACR-HR)

In this method, a judgment is made for test sequences where they are presented one by one and independently evaluated on a scale. This test should include a reference version of each test sequence presented like any other test stimulus. The method specifies that, after each presentation, observers are asked to assess the quality of the sequence shown. This method should be used in conjunction with a reference video where an expert in the field has rated it as good or excellent quality. It is also unable to analyze the degradation of the video. that occur in the first and last seconds of the video sequence.

The Double-Stimulus Continuous Quality-Scale method (DSCQS)

Also called double-stimulus method, it implies that the test sequences are presented in pairs: the first sequence of each pair is always the reference, while the second is the same but using its degraded version.

Degradation category rating (DCR)

This comparative method works as follows; the test sequences are presented in pairs, the sequences are presented on a first test system and then on a second one which is totally different. The systems under test (A, B, C,...,etc.) are usually tested by combining the different possibilities AB, BA, CA,.... . After the presentation of each pair, a judgement is made on the most preferred stimulus.

Pair comparison method (PC)

In order to have the video database that represents the best the visual quality perceived by our SVH, several factors must be taken into consideration. These include the viewing conditions, the choice of observers, the images or videos used, etc... . For a better understanding of all these factors, the reader is directed to the recommendations given by the VQEG group. The test results should be given with all the details of the experiment. For each test sequence the mean opinion score (Mean opinion score MOS) should be given. It will be used for the evaluation of objective metrics.

5.2.1.2 Objective quality evaluation

Subjective assessment remains the best way to estimate visual quality. The problem with such methods is the cost in terms of time and resources. This has led researchers to think of other assessment methods that are possibly objective metrics. With this kind of methods, the evaluation of perceptual quality is done automatically using algorithms. They use for their design the mathematical characteristics of the image that are likely to be affected by the artifacts. The objective evaluation metrics are divided into three groups: full-reference, reduced-reference, and no-reference metrics.

Full-reference metrics

The algorithm has access to a perfect version of the image/video with which it can compare the degraded version. The perfect version usually comes from a high quality acquisition device, after which it is degraded by compression and transmission errors.

Reduced-reference metrics

For this category, just a partial information about the perfect version is available. A channel side exists by which any information about the reference may be made available to the quality assessment algorithm. Reduced reference algorithms use this partial reference information to judge the quality of the distorted image/video.

No-reference metrics

the algorithm only has access to the distorted image/video and must estimate the signal quality without knowledge of the perfect (reference) version. Since non-reference methods do not require any reference information, they can be used in any application where quality measurement is required.

Any proposed new metrics must be evaluated for effectiveness. Their outputs should be correlated with subjective scores in a predictable and repeatable manner. Linearity between the two objective and subjective scores is not crucial. The most important and critical determinant of performance is the stability of the relationship. To remove any non-linearity due to the subjective assessment and to facilitate the comparison of scores, performance will be estimated using a non-linear regression between the set of objective and subjective scores [168].

After performing the linear transformation, we obtain the average MOS_p opinion scores, which are the predicted scores. Using the parameters listed below, the performance of the metrics can be evaluated.

$$MOS_p = \frac{1}{N}\sum_{i=1}^{N} score_{pi} \tag{5.1}$$

- **Pearson Linear Correlation Coefficient (PLCC)** : It is the co-variance between subjective MOS and objective MOS_p.

$$PLCC = \frac{\sum_{i=1}^{N}(MOS_{pi} - M\bar{O}S_p)(MOS_i - M\bar{O}S)}{\sqrt{(MOS_{pi} - M\bar{O}S_p)^2}(DMOS_i - DM\bar{O}S)^2} \tag{5.2}$$

where N is the number of images used. MOS_p, MOS, $M\bar{O}S_p$ and $M\bar{O}S$ are the objective score, the subjective score and their average values respectively.

- **Spearman Rank Order Correlation (SROCC)** : It calculates the order in which the values of the two vectors MOS_p and MOS appear.

$$SROCC = 1 - \frac{6}{N(N^2-1)}\sum_{i=1}^{N}(d_i^2) \tag{5.3}$$

where d_i is the difference between the ranks of the objective score and the subjective score of the tested image/frame.

- **Root Mean Square Error (RMSE)**: This is the root of the square mean of the differences between the subjective MOS and objective MOS_p scores.

$$RMSE = \sqrt{\sum_{i=1}^{N}(MOS_{pi} - MOS_i)^2} \tag{5.4}$$

MOS_p and MOS are the objective scores after fusion and the subjective scores.

5.3 The most common video surveillance distortions

Despite the huge progress that has been made towards the development of smart efficient security systems, the existing solutions present some limitation specially in cluttered and complex environments such as crowded places like in the case of a busy football stadium or highways traffic [169]. In such environments, captured video is subject to a wide range of degradation and distortions resulting from the different stages of the communication process. In fact, video distortions are generally classified into four different types based on their origin in the communication process starting from the capture, the compression, the delivery to the display. the most common distortions for video surveillance systems are listed below:

- **Coding artifacts** : coding artifacts are the most significant types of distortion. By reducing the temporal and spacial resolution it affects the visual video quality and it become more visible when the compression ratio increases.

- **Noise** : surveillance video may be affected by noise during acquisition due to camera sensors.

- **Blur** : Two types of blur can occur during surveillance process: Blur caused by out-of-focus is a very common problem and can affect the performance of surveillance operation, and blur caused by object/camera motion which may degrade the visual quality of the video.

- **Uneven illumination**: Surveillance videos are often affected by uneven illumination where certain parts of the video are bright, being within the sun light, while the others are in the dark.

- **Smoke**: For outdoor environment, smoke is a common distortion that may affect the video quality. In fact, a near fire can provides a lot of smoke in the camera's light of sight leading to a hard visibility .

- **Haze**: In some regions, haze is a very common natural phenomenal. It often occurs when dust and smoke particles accumulate in relatively dry air. In the presence of the haze, outdoor video is degraded by suspended particles making the far landscapes and objects invisible or hard to be seen. In some dried regions, another type oh haze can occurs in dried regions called heat haze which is an atmospheric condition that occurs when a body such as the ground is reflecting a lot of heat. It is caused by the difference in temperature between the hot body and cooler air around it.

In our dataset we will focus only on the four most important distortions for video surveillance context which are: Blur, Noise, Uneven illumination and Smoke.

5.4 Dataset creation

We constructed a new database containing 14 color full-HD (1920,1080 pixels) videos and 210 distorted videos. The database is built with free-use youtube videos from live streaming and some common videos used by the video processing community.

Youtube video are provided by a video surveillance equipment's provider from Netherlands called WebCam.NL [170]

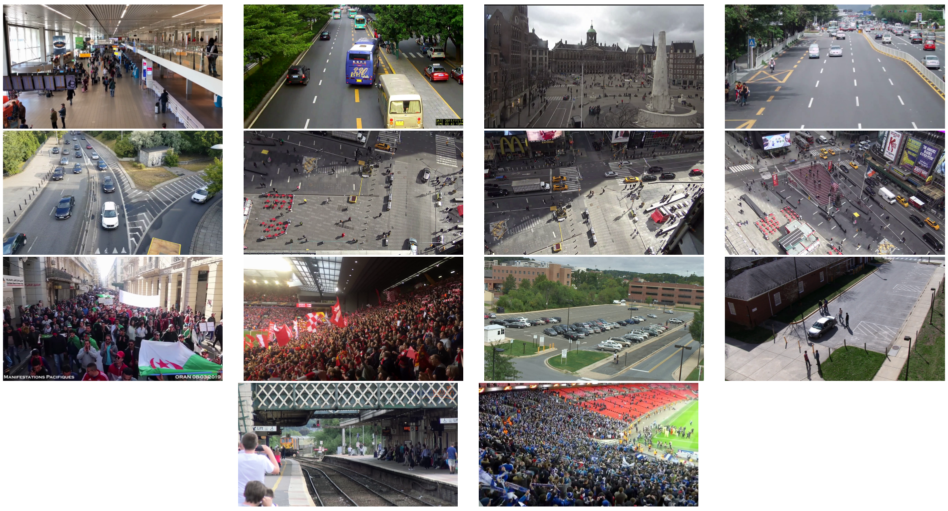


Figure 5.1: *One frame from each of the reference videos in the database.*

5.4.1 Distortions levels

We apply four different levels of the different distortions on the 14 original videos. This number of level is calculated in order to give more accuracy and efficiency to our database. In fact, In the following we present some figures of the distortions levels.

5.4.2 Distortions generation

As we said before, We have chosen four different distortions for our database. These distortions, which are among the most common affecting Visual surveillance videos, are the smoke, noise, uneven illumination and blur. In order to simulate each of this distortions, we have applied model for each distortion to each frame of the video. The details of modeling of each of these distortions are given below.

Additive White Gaussian Noise

surveillance video may be affected by noise during acquisition due to camera sensors.. Such noise can be modeled as additive white Gaussian noise as we have done for our database using the equation below for each frame $i(x,y)$.

$$d(x,y) = i(x,y) + N(x,y), \tag{5.5}$$

where $d(x,y)$ is the resulting distorted frame and $N(x,y)$ is the normally distributed random signal with the following probability density function

$$p_N(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\bar{z})^2}{2\sigma^2}} \tag{5.6}$$

In order to create four different levels, the variance of the Gaussian model was varied.

Uneven Illumination

Surveillance videos are often affected by uneven illumination where certain parts of the video are bright, being within the sun light, while the others are in the dark. In order to simulate this kind of distortion, we have initially generated a greyscale circular mask $m(x,y)$ having a bright circular region in the center and fading intensity towards the corners using the set of equations below

$$m(x,y) = \left\{ \begin{array}{ll} 1, & \text{for } d \leq h/AF \\ a, & \text{for } d \geq 2h/AF \\ 1 - \frac{(1-a)(d-h/AF)}{(h/AF)}, & \text{otherwise} \end{array} \right\} \tag{5.7}$$

where h is the height of the image, AF is the area factor, a is the attenuation and d is the distance of point (x,y) from the center of the circular regions. In real scenarios with uneven illumination, the bright region is often not in the center of the frame but rather pointed to one of the sides. For this reason, we have chosen the center of the mask to be slightly sideways to create a realistic effect. Using this mask, uneven illumination is then added to the original frame $f(x,y)$ using simple pixel-by-pixel multiplication to give the distorted frame $d(x,y)$

$$d(x,y) = C(x,y)f(x,y), \tag{5.8}$$

In order to create four levels of this distortion, we modify both the center point of the mask and the area of the central region.

Blur

Due to frequent and random motion occurring in the captured scene during the video surveillance process, often times blurriness due to motion can be observed, causing visual discomfort to the observer. This distortion was generated using motion filter implementation of MATLAB. We have simulated horizontal blur only and varied the length of the filter to generate different levels of distortion.

Smoke

For outdoor environment, smoke is a common distortion that may affect the video quality. In fact, a near fire can provides a lot of smoke in the camera's light of sight leading to a hard visibility. It is a difficult distortion to synthesize for videos. In order to generate videos with smoke, we have made use of a video containing only the smoke with a black/Green background. This video is then blended with the reference video using screen blending mode to create the effect of smoke in the video.

$$v(x,y) = 1-(1-r(x,y))(1-\alpha s(x,y)), \tag{5.9}$$

where $r(x,y)$ and $v(x,y)$ are the original and the resulting distorted frames respectively, α is the opacity and $s(x,y)$ is the smoke video frame. By adjusting the value of opacity α we generate different levels of smoke in the video.

5.4.3 Video categories

Our videos are divided into four categories and they are all day-captured videos. These categories present the most common places where video surveillance systems are the most used:

- **Crowded Street** : Since the very first use of video surveillance cameras, crowded streets and roads were the most important place to be deployed in.

- **Transport sites** : Transport places like railways, metro, train stations and airports are sensitive and high-risk sites and smart video surveillance systems are highly recommended to ensure safety and security.

- **Parking** : parking sites are among the most risky places that need to be supervised and monitored by smart video surveillance systems.

- **Stadium** : stadiums are the most crowded sites where security consists a nightmare for officials. thus, smart video surveillance system with intelligent action/face recognition can helps reducing the risk probability.

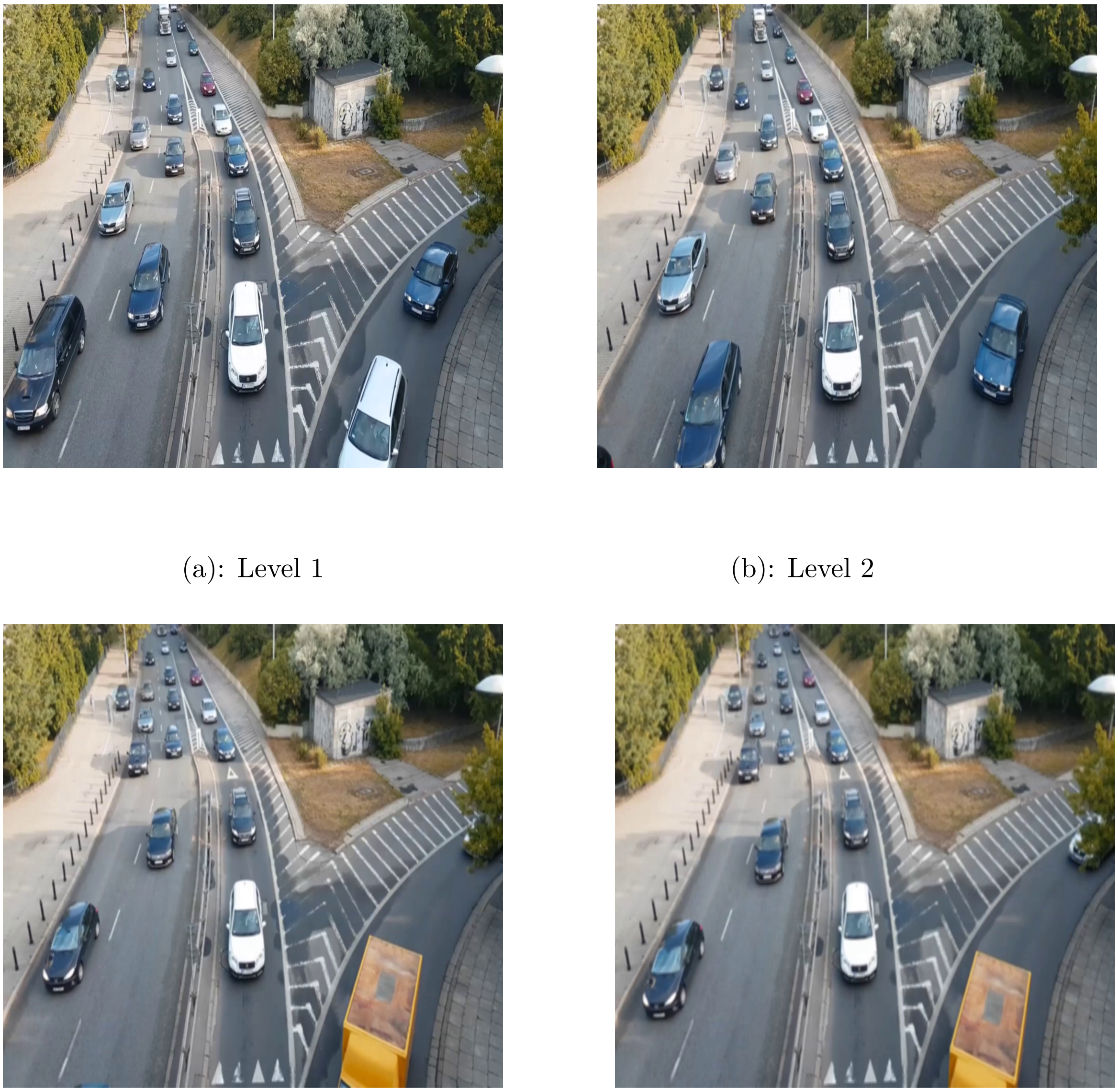
5.4.4 Distortions levels

We apply four different levels of the different distortions on the 14 original videos. This number of levels is calculated in order to give more accuracy and efficiency to our database. In the following we present some figures of the distortions levels.

5.4.5 Video Quality Score and performance

Once the distortion is detected and identified in a video, it is possible to evaluate its severity using a quality score. To this end, we have selected 2 different metrics which are often used as benchmarks for natural images and videos. These are PSNR and SSIM. We have also used a full-reference (FR) metric VIF [171]. However, for video-surveillance videos, there is usually no ground truth available and a no-reference (NR) metric makes more sense. Therefore, we have also included No-reference metrics. Thus, We have included two NR image quality metrics BRISQUE [172] and NIQE [173]. For both of these, we have used the mean metric value from all frames as the score for the video.

A total of 19 observers carried out the subjective tests for the database. Outliers were first identified and detected on the basis of non-transitivity. The preference matrices for the remaining subjects were then compiled and aggregated to obtain subjective scores. Furthermore, in order to evaluate whether an existing objective video quality metric correlates well or not with subjective scores, Spearman Rank Order Correlation (SROCC) and Pearson Linear Correlation Coefficient (PLCC) were evaluated for the metric scores after performing a non-linear regression with a 5-parameter logistic function. From Tables 5.1 and 5.2, we can see that none of the objective metrics perform well when overall correlations are evaluated, with maximum PLCC value of 0.5379 with BRISQUE. However, with individual



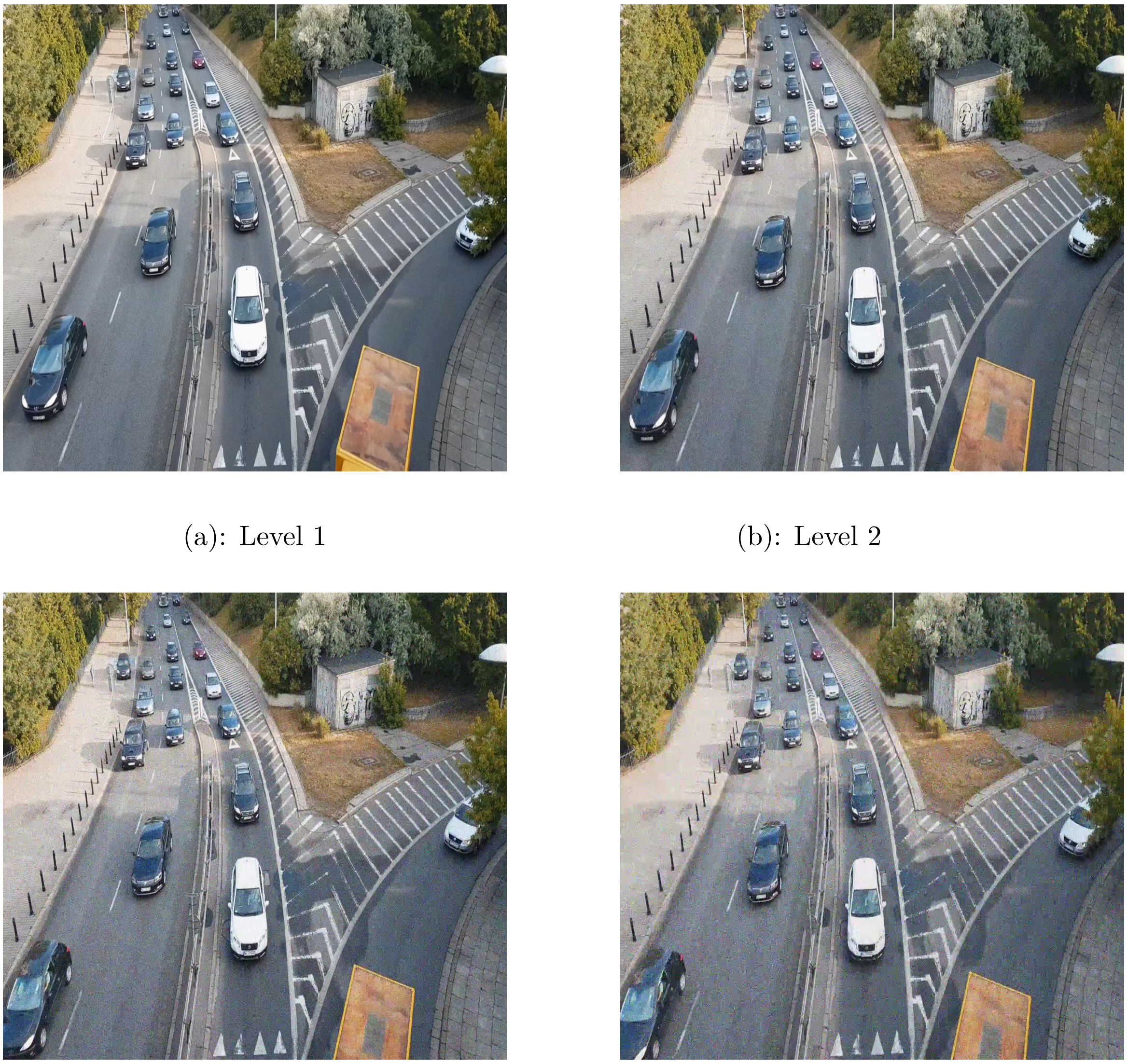
(a): Level 1

(b): Level 2

(c): Level 3

(d): Level 4

Figure 5.2: *Blur Levels*



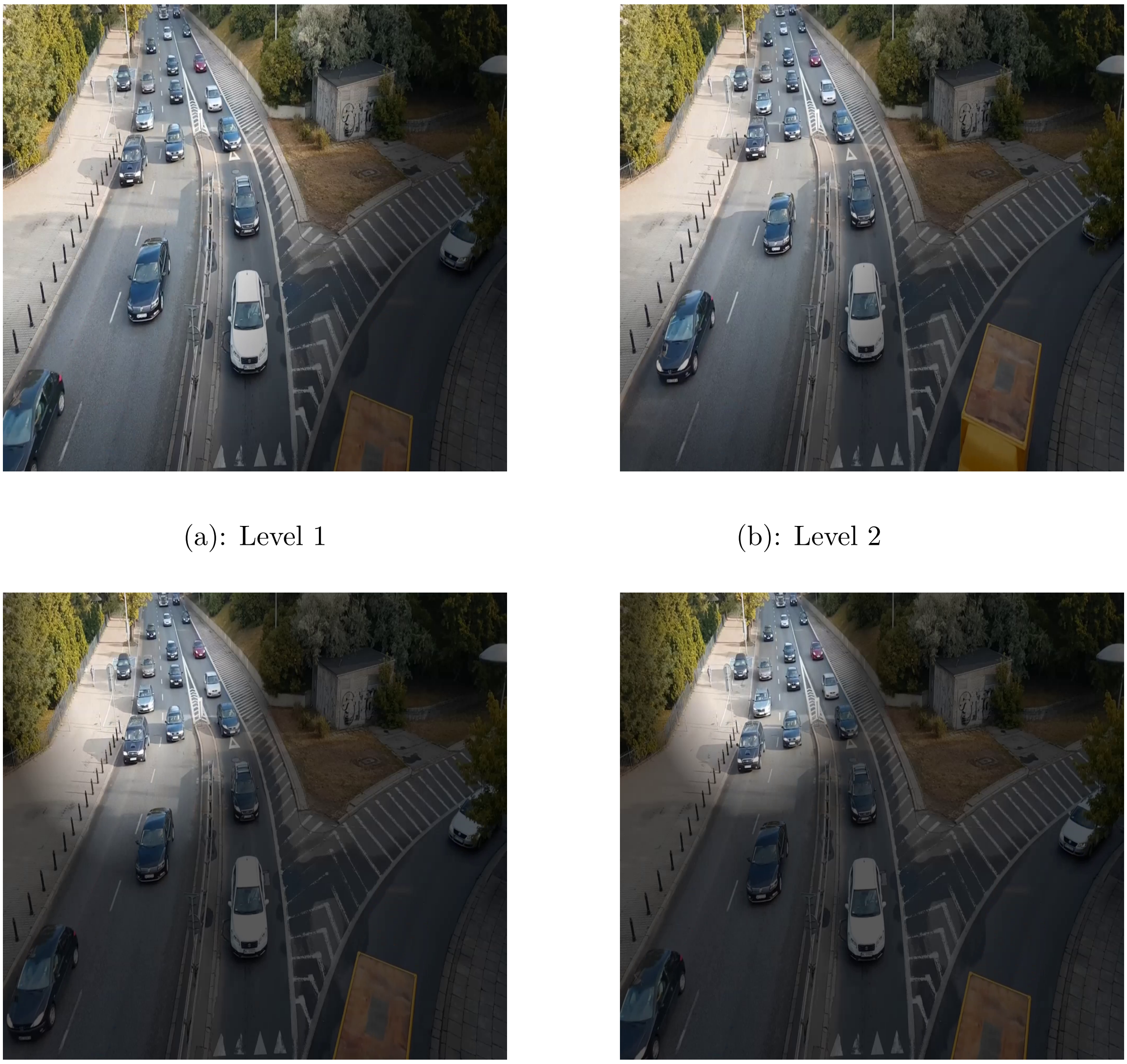
(a): Level 1

(b): Level 2

(c): Level 3

(d): Level 4

Figure 5.3: *Noise Levels*



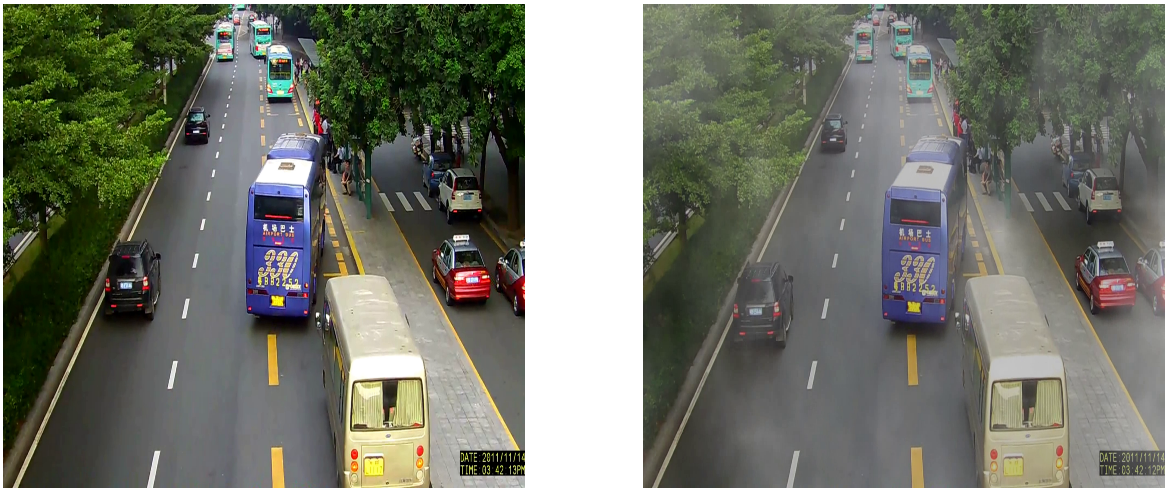
(a): Level 1

(b): Level 2

(c): Level 3

(d): Level 4

Figure 5.4: *Uneven Illumination Levels*



(a): Original video

(b): Smoked Video

Figure 5.5: *Smoke distortion*

distortion types, PSNR correlates much better with subjective scores as compared to others for both groups for uneven illumination and smoke distortions. Among the NR metrics, both NIQE and BRISQUE give acceptable results for blur, with BRISQUE being better for noise. We can conclude also that the no-reference metrics with the VIF one give worse results for Uneven illumination metric.

All these results are very significant as they imply that none of these metrics are generic or non-distortion specific for the kind of videos and distortions encountered in video surveillance context. To this end, it is very important to develop a video-surveillance oriented objective metrics that can perform well when overall correlations are evaluated.

Table 5.1: *PLCC scores in Database (best value in bold for each column)*

Metric	***UI***	***Noise***	***Blur***	***Smoke***	***Overall***
PSNR	**0.8424**	0.3784	0.3066	**0.9568**	0.5029
SSIM	0.7701	0.3784	0.3066	0.8558	0.4786
BRISQUE	0.4297	**0.7397**	0.7930	0.6101	**0.5356**
NIQE	0.2985	0.6611	**0.8687**	0.4940	0.4552
VIF	0.2999	0.6652	0.6845	0.8200	0.4844

Table 5.2: *SROCC scores in Database (best value in bold for each column)*

Metric	***UI***	***Noise***	***Blur***	***Smoke***	***Overall***
PSNR	**0.7327**	0.3744	0.1949	**0.9305**	0.3910
SSIM	0.6062	0.6377	0.4846	0.8240	*0.4367*
BRISQUE	0.3713	**0.7266**	0.7766	0.5993	**0.5379**
NIQE	0.2458	0.6841	**0.8475**	0.4441	0.4414
VIF	0.3030	0.6104	0.6238	0.7642	0.4630

Study of metrics monotonicity

In order to study the behaviour of some of the presented metrics on the proposed distortions, we have created a video for every distortion from the same reference video with progressive increased distortion severity (10 seconds video). In fact, we have chosen to study the uneven illumination case For the No-reference metrics BRISQUE and NIQE we have plot the scores for some frames in the progressive distorted video and its correspondence in the reference one. It is important to mention that the distortion severity changes every 10 frames for blur and noise, 20 frames for the uneven illumination and every 30 frames for the smoke.

The figures 5.6, 5.7, 5.8 and 5.9 draw the curves of BRISQUE metric. We can see that BRISQUE score is slowly increasing when the noise intensity increase which means a quality diminution. Therefore, we can conclude that BRISQUE can be used for noise detection in video but it is not high sensitive its evolution in time. For blur, we can see that for the first four points (80 frames) BRISQUE score is increased very fast which means an efficient detection of the blur severity evolution. However, starting from the 100th frame we see that the curve is stabilized which means that the BRISQUE is not detecting the high change in the blur distortion rate. Therefore, BRISQUE is not a suitable metric for blur. For smoke distortion, we can see that the BRISQUE score is increasing when the smoke severity increases which mean that this metric is able to detect the smoke rate changes in the video. However, the metric behaviour can not be understood for first 60 frames where it gives lower scores for distorted video compared to those given to the original video

which means that accordingly to BRISQUE for these frames, the distorted video has better quality than the original one. Surprisingly for uneven illumination distortion, BRISQUE is given for all distorted frames lower score than those given to the original ones which means that the distorted video presents better quality compared to the original one. Therefore, we can conclude that BRISQUE is totally unsuitable for uneven illumination distortion. In fact, this conclusion proves the results found in the SROCC and PLCC tables.
The figures 5.10, 5.11, 5.12 and 5.13 draws the NIQE results for the same context. For the noise, we can see that for a long period of the video (from 60th frame to the 180th one) NIQE did not detect well the noise severity change. Then for the last frames, NIQE behaviour is not comprehensive with some oscillations which means that NIQE is not suitable and appropriate for noise distortion. For blur, we see that NIQE is able to detect well the severity change. Therefore, we can judge that NIQE can be an appropriate metric for blur. NIQE behaviour for smoke is similar to the BRISQUE one except the first frames where NIQE is able to detect a quality degradation for the distorted video compared to the original one. Therefore, NIQE would be a suitable metric for smoke distortion. Finally, we see that for uneven illumination distortion NIQE presents the same behaviour as the BRISQUE one for the same distortion which means that NIQE is not totally not suitable for uneven illumination distortion.

We can conclude that BRISQUE curve is oscillating and its not stable comparing to the curves of PSNR and SSIM which are relatively stable. Otherwise, VIF is not stable and it presents the worst curve. Surprisingly, the results show that sophisticated metrics such as BRISQUE are not stable for our Uneven Illumination video comparing to the oldest ones; PSNR and SSIM. We know that the content of video frames changes over time. This may produces some oscillating behaviour in the metric. In fact,the light will interact with the content of the scene and the reflected light depends on the physical and optical properties of the materials in the scene. So if the content of the scene changes over time, the frames may be totally different which lead in certain circumstances to increase or decrease the uneven illumination effect in the scene. However, even if such thing may happen,

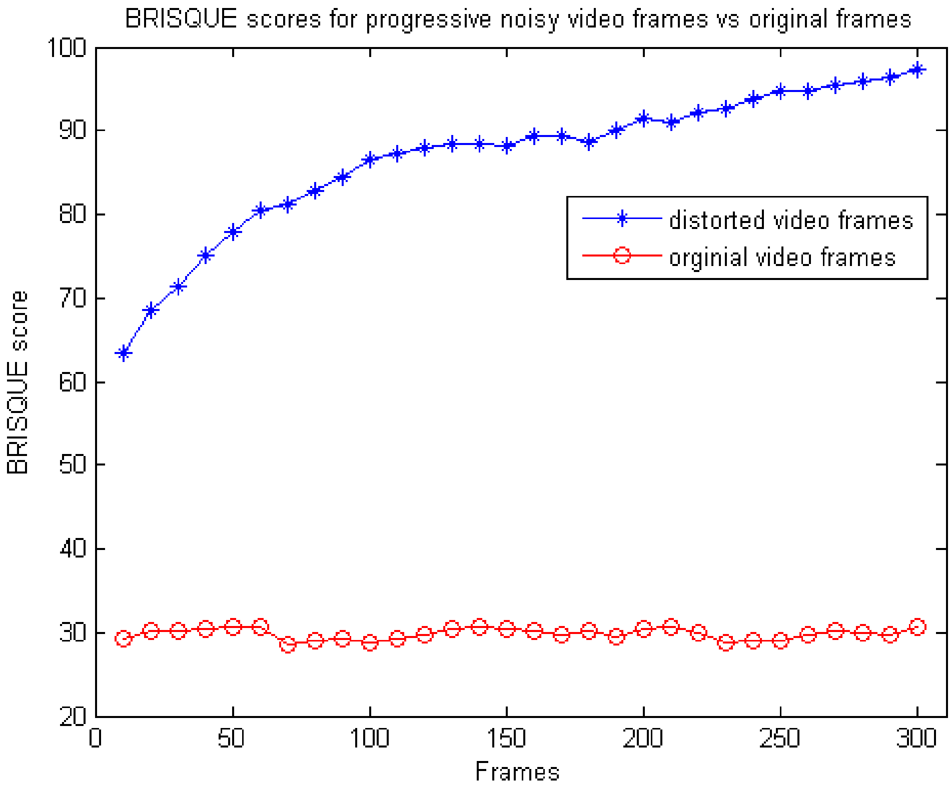


Figure 5.6: *BRISQUE metric monotonocity for Noise distortion*

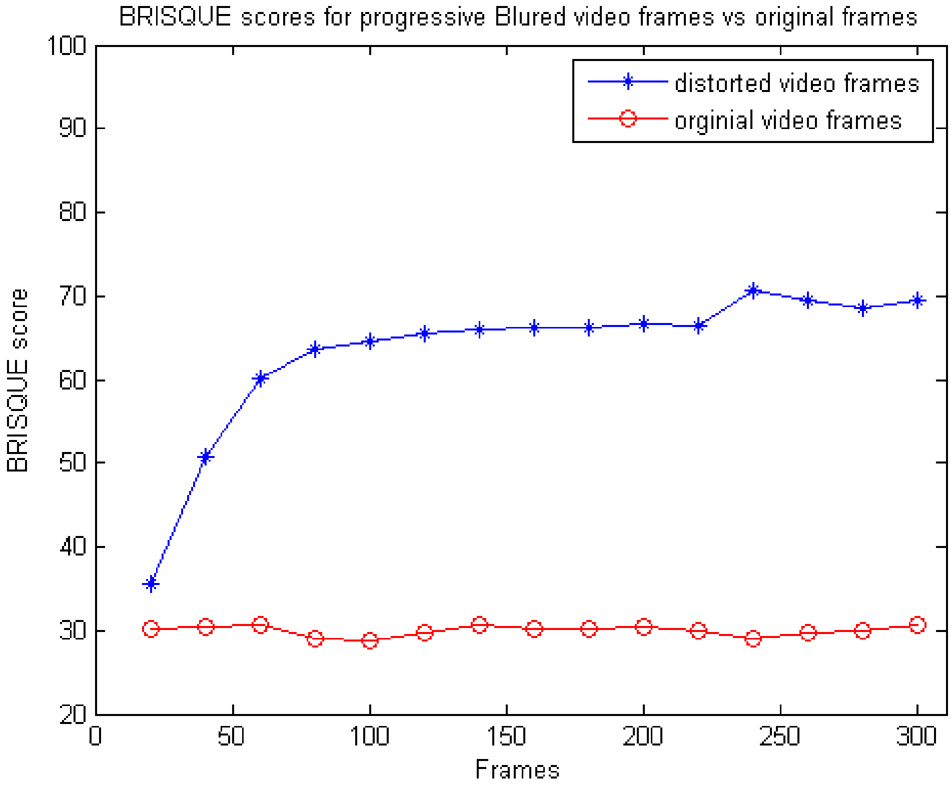


Figure 5.7: *BRISQUE metric monotonocity for Blur distortion*

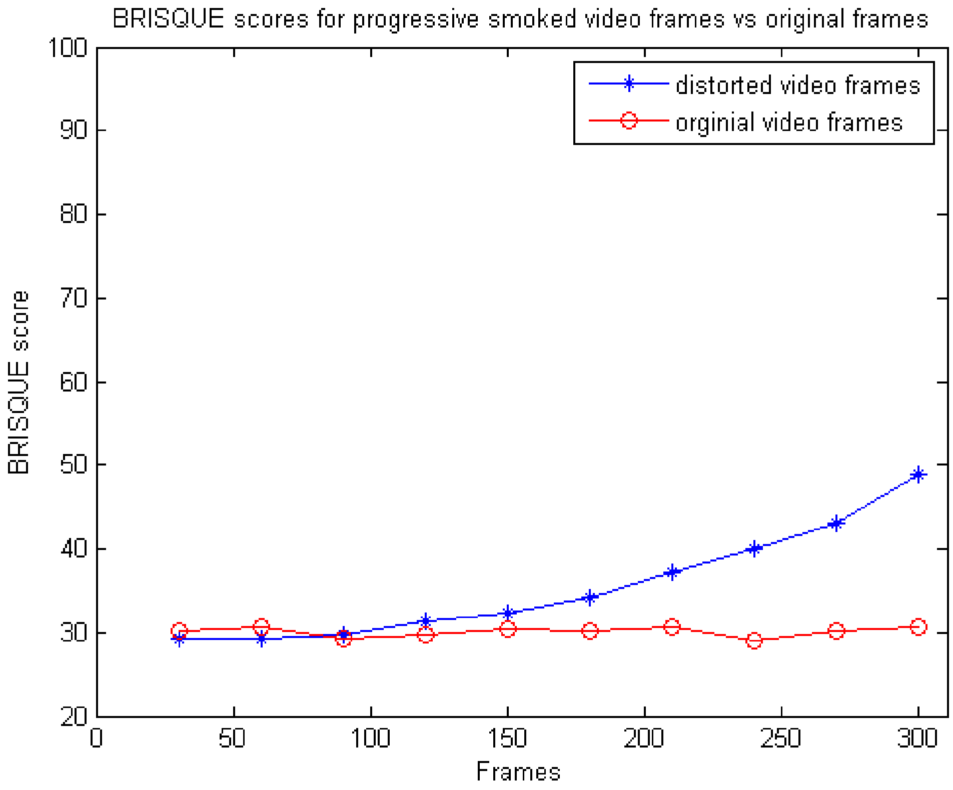


Figure 5.8: *BRISQUE metric monotonocity for smoke distortion*

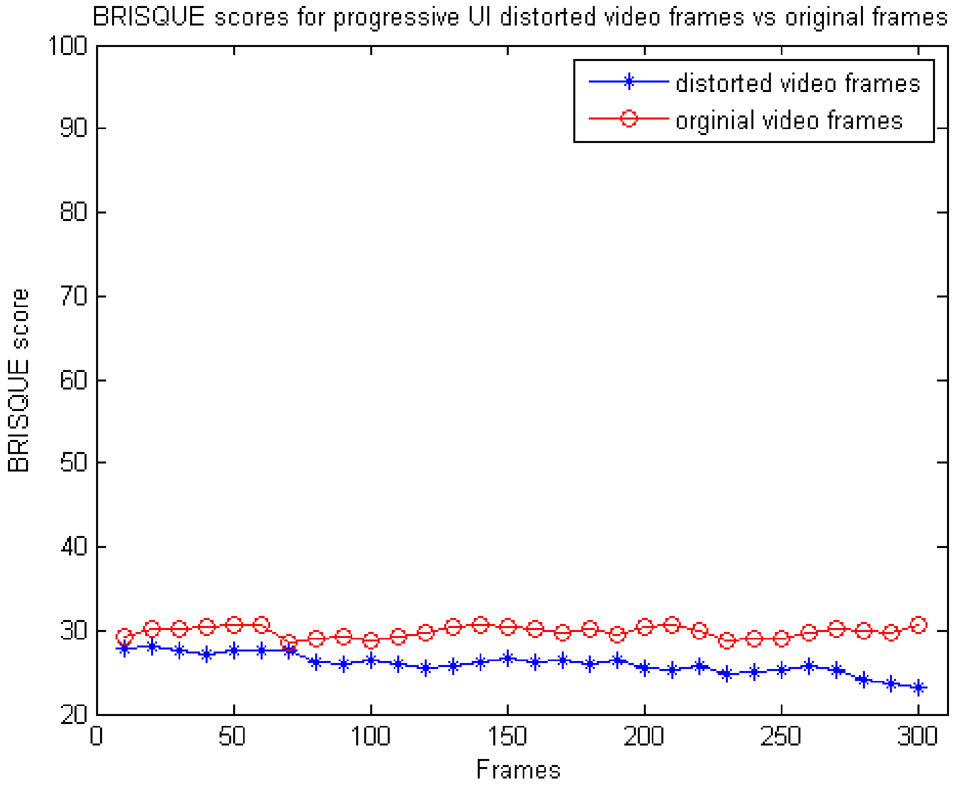


Figure 5.9: *BRISQUE metric monotonocity for Uneven Illumination distortion*

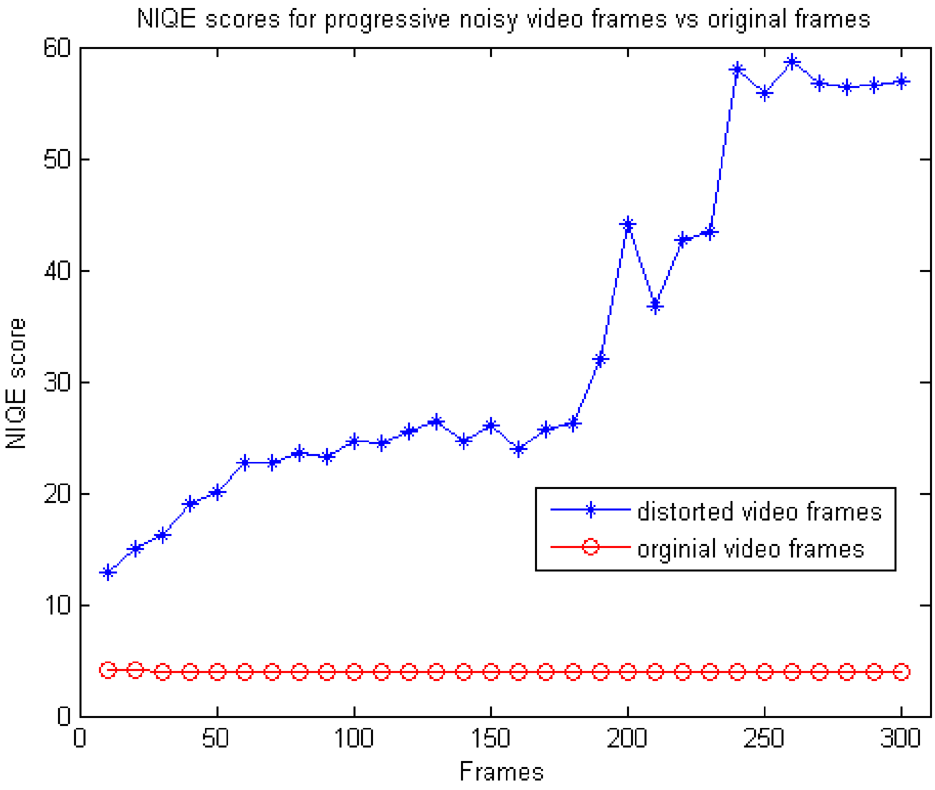


Figure 5.10: *NIQE metric monotonocity for Noise distortion*

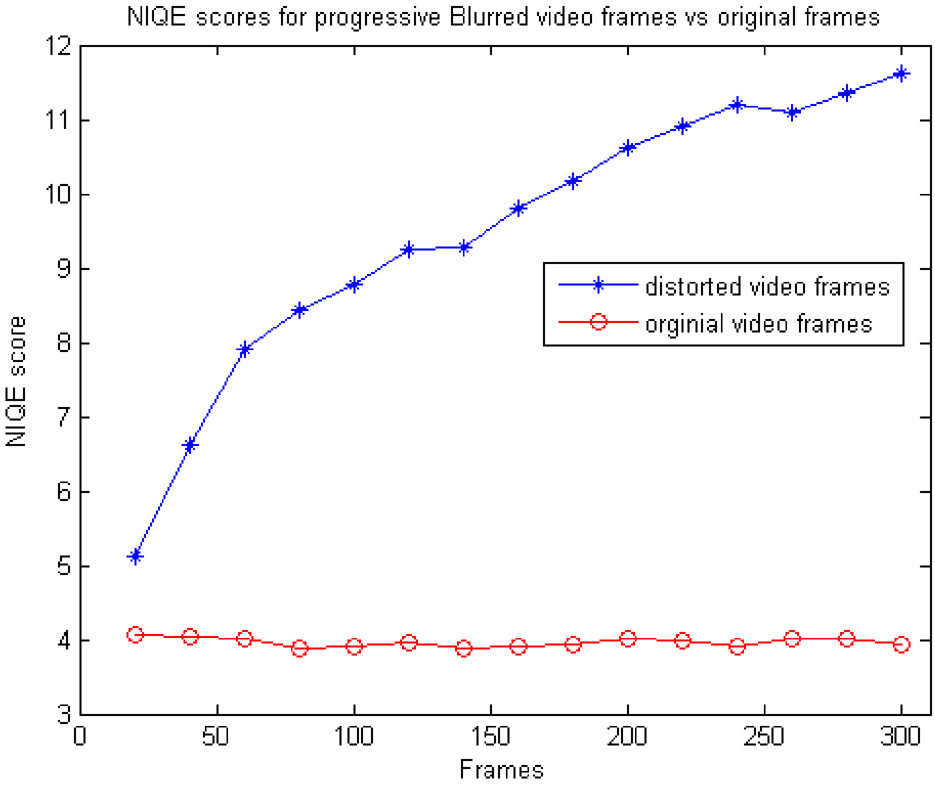


Figure 5.11: *NIQE metric monotonocity for Blur distortion*

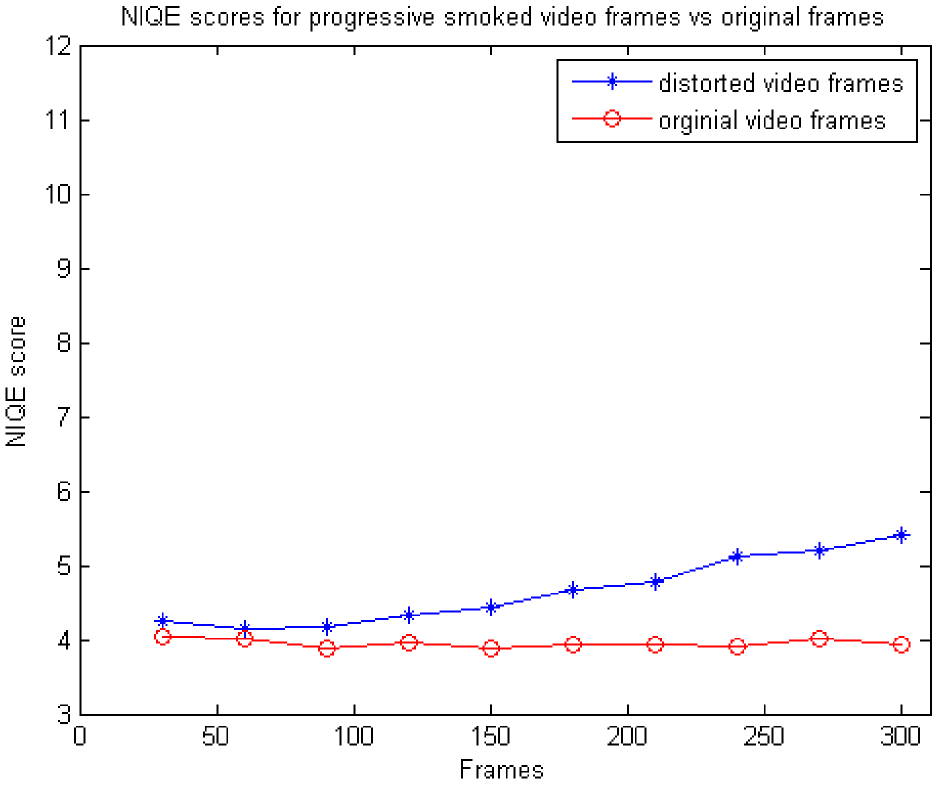


Figure 5.12: *NIQE metric monotonocity for smoke distortion*

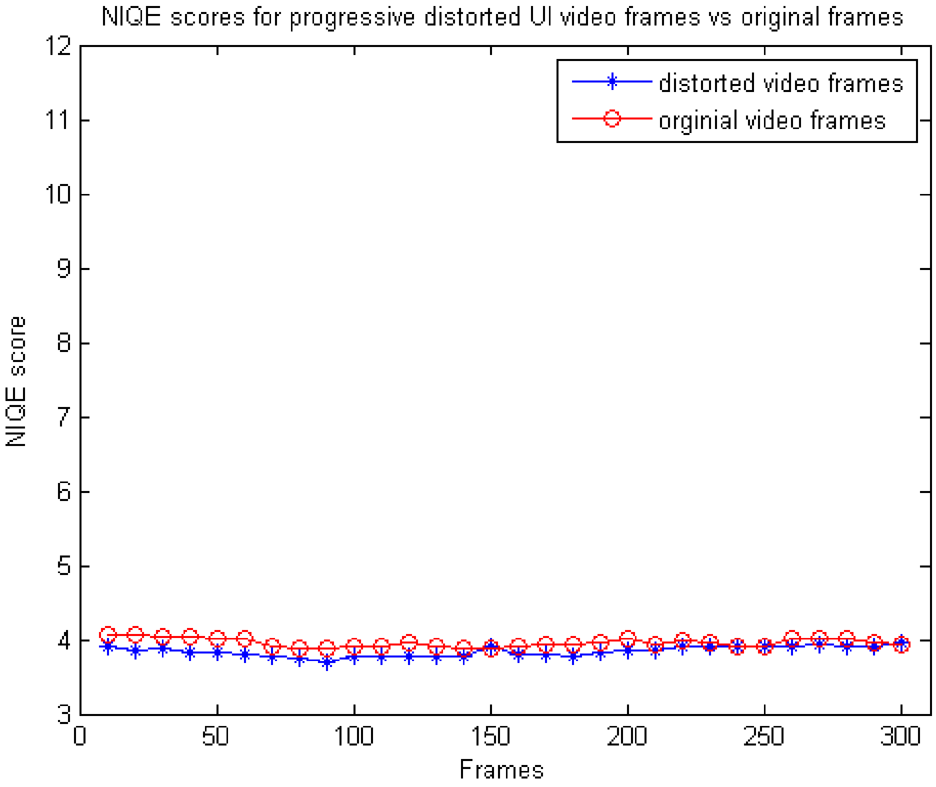


Figure 5.13: *NIQE metric monotonocity for Uneven Illumination distortion*

it would not produce such highly irregular and oscillating behaviour in the metric and especially in the case of BRISQUE.

5.5 Towards a driven-quality smart surveillance system

Despite the huge progress that has been made towards the development of smart efficient video surveillance systems, the existing solutions present some limitation specially in cluttered and complex environments such as crowded places like in the case of a busy football stadium or highways traffic [169]. In such environments, captured video is subject to a wide range of degradation and distortions resulting from the different stages of the communication process. These distortions have been discussing in the previous section. In general, smart video surveillance systems rely on video analysis algorithms for automated video processing. However, little is known about the minimum video quality required to ensure an efficient performance of these algorithms. In [166], authors have concluded that the algorithms show almost no decrease in accuracy and efficiency until a certain critical quality of the input video is reached, which amounts to significantly lower bit-rate compared to the quality commonly acceptable for human vision. Authors in [167], have shown the effect of bad/uneven illumination on the face detection process. This illumination issue may degrade the perceptual video quality and affect the efficiency of the surveillance process.
In [167], authors demonstrated through their obtained results the importance of image/video quality enhancement in the context of video surveillance. However, to enhance the quality we need to asses it before, in order to have a correct evaluation of the video quality.Therefore, it is very important to include a video quality assessment box in the video surveillance system architecture in order to evaluate the video quality ,thanks to some specific metrics, and make a decision on an eventual quality enhancement.

Since the beginning, video quality has not been seen as an important factor in video surveillance. Indeed, conventional systems and even those with a

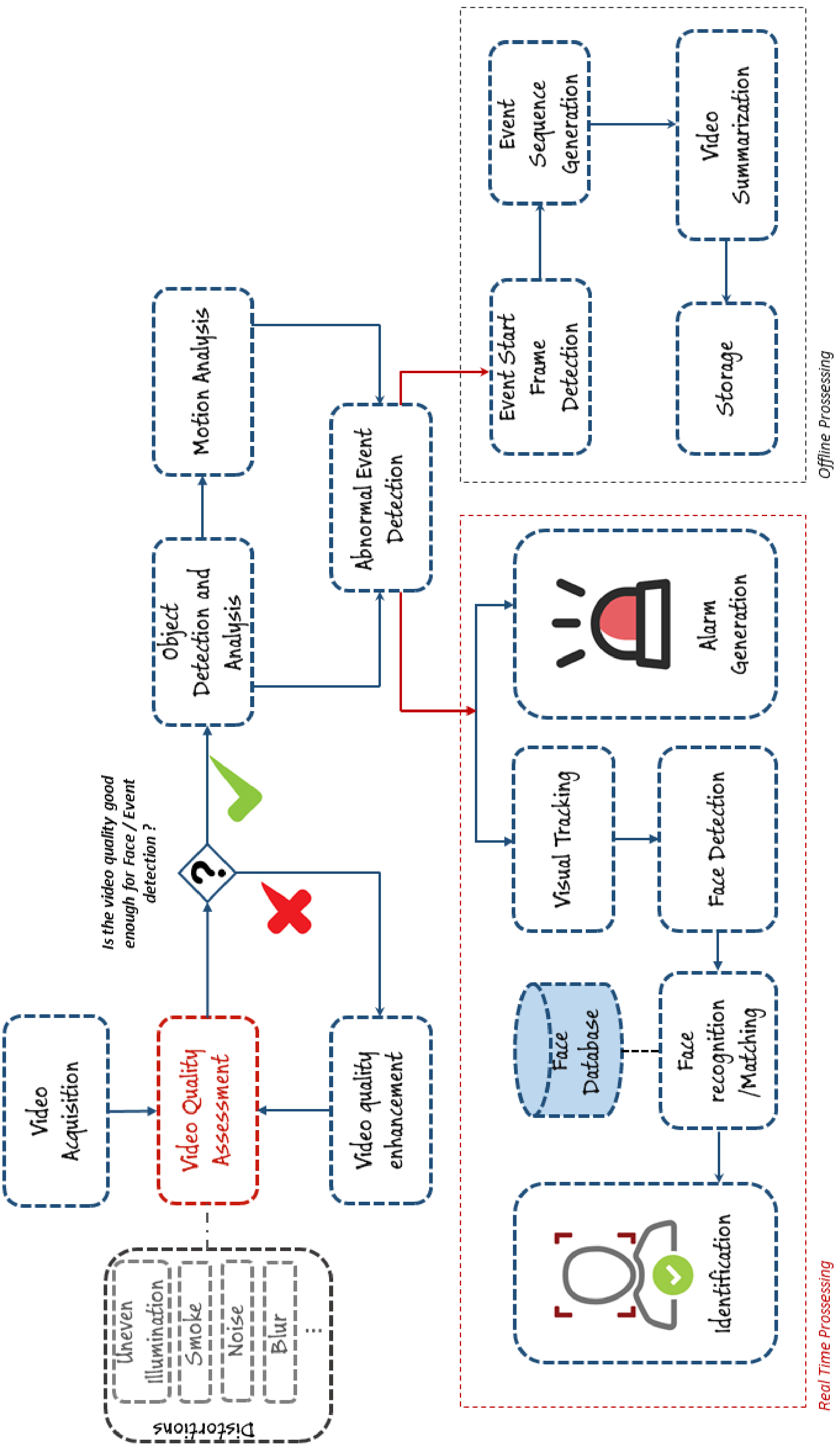


Figure 5.14: *Quality-based intelligent video surveillance architecture*

minimum of intelligence ensure an acceptable visual quality for the monitor to identify abnormal events. However, when the suspect identification is necessary, it could be done offline where a quality assessment and enhancement process can be performed if we cannot get a good detection and classification of faces or actions in the video. In some cases, the detection of objects or/and events and the identification of suspects must be done in real time to allow efficient decision making. Thus, we propose in this section a driven quality smart video surveillance architecture. Fig. 5.14 shows all the blocks of the proposed video surveillance architecture.

5.5.1 Smart video quality assessment box

After video acquisition, VQA box will decide based on some metrics scores if the video quality needs be enhanced or not [174] [159]. In fact, one of the most challenging concern for this quality-based video surveillance systems is the distortions detection, identification and classification. Therefore, surveillance video can be affected by more that one distortion which may lead to an unimproved quality after enhancement step based on the scores given by a single-distortions quality assessment metrics. For this reason, it is primary to think about a smart system for distortions detection, identification and classification:

- **Distortions detection** : In some cases the surveillance video can be affected by more that one distortion. For example, uneven illumination with smoke would be difficult to be both detectable in the video bagsing on video quality assessment metrics for uneven illumination or smoke separately. We need to propose a leaning algorithms that can detect the presence of every distortion in the surveillance video.

- **Distortions classification** : After being detected, distortions need to be classified based on its intensity. For example a just noticeable smoke can be merged with a very annoying uneven illumination distortions. In this case, it's more prior to proceed to uneven illumination enhancement rather that smoke enhancement. Therefore, the smart system needs to be able to classify the level of each presented distortions

- **Distortions identification** : taking back the last example of a video with a just noticeable smoke and a very annoying uneven illumination distortions. The smoke distortion in this case is present but it may not affect the visual detection and tracking process and it will be need to proceed to a smoke enhancement and the system can just enhance the uneven illumination and proceed to the face detection/recognition and visual tracking steps. Thus, distortions identification means to choose which distortion needs to be enhanced based on the distortions classification made before.

The smart system for distortions detection, classification and identification identifies the most annoying distortions that can lead to a bad video quality and send this information to the quality enhancement box which will execute the most appropriate enhancement techniques depending on the distortion and its intensity.

Fig. 5.15 present the smart Video quality assessment and enhancement system.

In literature, distortion features of video can be used to estimate the video quality. To estimate the perceived visual distortion in video frames, structural distortion is employed in [175]. In [176] a no-reference method of video quality assessment uses a statistical distortion features. In fact, each frame is represented in the wavelet domain and the oriented band-pass response is generated by their decomposition [177]. The resulted sub-band coefficients are extracted with the statistical distortion features in order to construct a feature vector which will serve as a descriptor of the overall distortion of the frame. In the wavelet domain, the video quality is accomplished by the classification of the feature vectors across frames and a score mapping. A motion-compensated method using block and motion vector serves to evaluate the temporal quality. Finally, the overall video

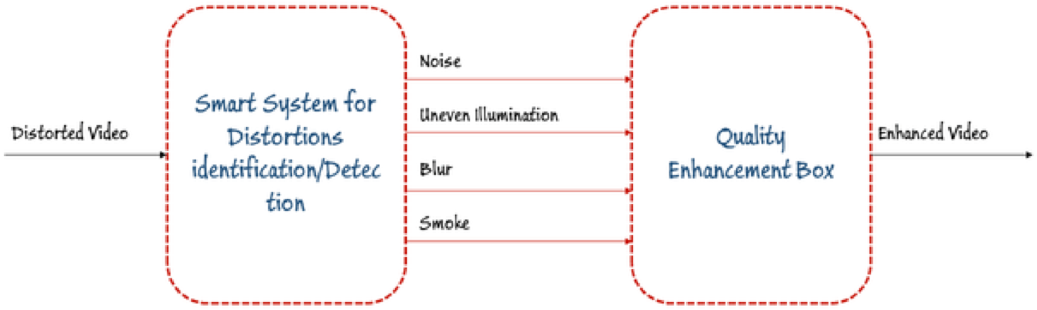


Figure 5.15: *smart video quality assessment and enhancement boxes*

quality is achieved by a pooling strategy.

5.5.2 Intelligent video processing: Abnormal event detection

Once the video quality judged good by the video quality assessment box, the system performs an object detection and analysis process in order to detect and classify every object in the scene under surveillance.

5.5.2.1 Object detection and motion analysis

Object detection is usually the basis of any intelligent video surveillance system. It allows the detection of activities in the monitored scene, such as the movement, appearance or disappearance of an object. Object detection aligns with motion detection since moving regions of the scene are of interest (foreground) and static parts are not (background) [12]. Many motion detection techniques are based on change detection. However, detecting changes in a scene may not necessarily leads to the movement of objects, but it can highlight a modulation of the image. In order to segment moving objects, we must be able to make difference between pixels corresponding to consistent motion and those caused by environmental changes. Complex environments can be a major problem because of the many variations (lighting changes, unnecessary movement, backgrounds, etc.) that can occur in complex and congested environments. In literature, the majority of motion detection methods are based on background subtraction technique. The later consists of two main steps: (1) Modeling the background; (2) Motion segmentation. Background modeling is the representation of the scene without the moving objects and must be updated regularly. Motion segmentation aims at detecting regions corresponding to moving objects (people, vehicles, ...). The frames of the video are compared to the background model and the differences are marked as moving objects.

Once moving objects are detected, their movements are tracked throughout the video sequence. Tracking is the estimation of the trajectory of an object in the image plane as it moves through the scene. This task requires locating each object from frame to another. Tracking can be done in 2D, from a single camera, or 3D, by combining two views with a known geometric relationship (This was one of the motivations that has pushed us to think about Multiview surveillance systems). In fact, Many tracking techniques predict the position of the object in a frame based on its movements observed in previous frames. Each detected object must be associated with its correspondent in the next frame to update its trajectory, otherwise a new trajectory is created. Tracking these objects can be difficult due to their complex shapes, their non-rigid nature, their movements, partial or complete occlusions, changes in scene lighting, etc. These can be simplified by simple assumptions such as smooth movements, and prior knowledge of the number, size, shape and appearance of the objects. Tracking allows the extraction of other characteristics: trajectory, speed, direction of movement, position at a specific time. In [55] [52] [53] [54], authors presents different classification of object tracking methods. The most popular classifications are those of [54] which classifies object tracking algorithms into four categories: region-based algorithms, active edge-based algorithms, feature-based algorithms and model-based algorithms.

5.5.2.2 Abnormal event detection

Beyond object/event information, a very important task in visual surveillance is to detect abnormal events [178]. It helps predict dangers and raise alerts to security staffs or police and as a result makes intelligent visual surveillance system valuable for end users. Therefore, Behavioural analysis is the highest level task performed by intelligent video surveillance systems. The information collected by the previous steps is interpreted through semantic description to describe the behaviours and interactions of objects in the scene with natural language. Semantic analysis is often highly dependent on the context of the application. The most commonly used techniques to model the detected behaviours are: Hidden Markov models [179], neural networks, Bayesian networks, Kernal discriminant analysis [180] etc. First of all, visual information of moving objects in the scene is extracted and described with an appropriate method, then these information are studied to recognize and understand behaviour and thereafter the event. Many features have been proposed to describe human activities based on three main algorithms [68]:

- **Algorithms based on 3D models** : The most common technique to reach the 3D information of a movement is to retrieve the pose of the person or object at each moment using a 3D model. The model is constructed by trying to minimize a residual measurement between the projected model and the contours of the object. This usually requires a strong foreground/background segmentation. As an example, Campbell and Bobick [69] who calculate the 3D information of the positions of human body limbs. Their system exploits redundancies that exist for particular actions and performs recognition using only the information that varies between actions. This method only examines the relevant parts of the body.

- **Algorithms based on appearance models** : Contrary to 3D algorithms, other works try to use only the 2D appearances of the action. An action is described by a sequence of 2D instances/positions of the object. Many methods require a standardized image of the object (usually without background). For example, authors in [70], [71], and [72] present results using actions (mainly hand gestures), where grayscale (backgroundless) images are used. Although hand appearances remain fairly similar in many people, with the obvious exception of skin colour, actions that include the appearance of the whole body are not as visually consistent in different people due to natural variations and different clothing appearances.

- **Algorithms based on motion models** : These approaches attempt to characterize movement without referring to static body poses. The authors of [73] use repetitive motion as a strong warning signal to recognize cyclic walking movements. They track and recognize people walking in outdoor scenes by collecting a vector that characterizes the whole body. This vector carries low-level movement characteristics and periodicity measurements. Other work, such as [74], focuses on movements associated with facial expressions using movement properties based on predefined regions. The goal of this research is to recognize human facial expressions as a dynamic system, where the movement of regions of interest is relevant. These approaches characterize the expressions using the properties of the underlying movements rather than representing the action as a sequence of poses.

After characterizing the behavior, its patterns are analyzed for recognition. For now, the recognized behaviours are mainly: head and limb movements and gestures [181]. There are two types of behaviour recognition algorithms, as follows [182]:

- **Template Matching method** : The basic idea is to extract characteristics from video sequences and then compare them with pre-recorded behavioral patterns. This method has a low computational cost, but it is sensitive to noise. Therefore, efficient noise enhancement method needs to be performed once a noise distortion is detected by the video quality assessment bloc.

- **State space method** : Each static gesture is defined as a state, then all these states are combined with a probability. Each behavior is considered as a set of states. The classification of the behavior depends on the maximum value of the joint probability. This method requires a complex iterative calculation.

In the last decade, some abnormal event detection learning-based methods have been proposed. Recent applications of convolutional neural networks have demonstrated promises of convolutional layers for object detection and recognition. Authors in [183] proposed an efficient method for detecting anomalies in videos based on convolutional neural networks. To this end, a spatiotemporal architecture was proposed for anomaly detection in videos including crowded scenes. In [184], authors address the abnormality detection problem in crowded scenes using generative adversarial nets (GANs).

5.5.3 Real time processing

Once an abnormal event is detected, the concerned persons must be tracked and identified in parallel with generating an alarm to prevent the security agent's and to facilitate their intervention. In surveillance systems for suspects searching [185], the identity of the suspect must be revealed, for example through facial recognition of the individual. A lot of research has been invested in this application in recent years. Facial recognition is one of the main tools used for biometric identification of people on video [186] because it allows more precise identification. However, face recognition in an uncontrolled environment remains a problem that has not yet been satisfactorily resolved.

5.5.3.1 Offline processing

While the abnormal event detection

visual surveillance is usually carried out 24 hours 7 days per week and hence can generate tremendous amount of video data. End users usually demand a simple user interface to allow them browse multi-sensor data in a friendly and quick way without losing key information. Usually, this done through the concept of video summarization and visualization [187],[188] where the most important information from videos is extracted and abstracted into a concise form using visualization techniques to allow users browse the vast amount of visual data in a very short time. Therefore, from the huge amount of visual information, the most relevant sequences will be the only ones to be extracted. In the case of video surveillance,The relevant sequences will be summed up in a short video containing only the video clips of the detected abnormal events. These summarized videos could be used for quick scene analysis and interpretation. This will be achieved by using the results derived from the previous tasks which is abnormal event detection.

5.6 Conclusion

In this chapter we presented the video surveillance database that we have created in order to help the scientific community working on video quality assessment and especially for video surveillance application. The second part of this chapter was dedicated to present our quality-based intelligent video surveillance architecture where we have proposed an innovative idea about video quality assessment smart box able to detect, classify and identify the distortions that may be present in the captured video. This task is very important in order to decide whether or not the video must be enhanced. A good video quality offers better scene understanding and an efficient event detection.

CHAPTER 6

Concluding remarks and perspectives

VIDEO surveillance is increasingly receiving a lot of attention as an active area of research. Intelligent video surveillance is a technology that automatically identifies specific objects, behaviours or attitudes in video footage. It transforms video into data that will be transmitted or archived to enable the video surveillance system to act accordingly. This may involve activating a mobile camera, in order to obtain more precise data from the scene or simply to send an alert to the surveillance personnel so that they can make a decision on the appropriate action to be taken. Recent sophisticated systems guarantee good and high object/face or event detection, recognition and tracking. Therefore, a good video quality is needed to ensure this process efficiently. However, in such system different signal processing stages can affect the acquired video quality; capture, network transmission, video compression, causing video distortions. Video surveillance systems are often deployed in outdoor structures; stadium, buildings, parking and streets but to name few. Based on the outdoor environment nature, some natural degradation due to adverse weather conditions such haze, fog and smoke in some cases greatly reduce the visual quality of outdoor surveillance videos and make the event/object detection and recognition process not efficient.

Therefore, any video quality degradation impacts directly the efficiency of the video surveillance system which may straightly affect security. The quality of experience of a video surveillance systems is highly dependent on the quality of the video but also on the quality of service of the network. In fact, captured video may present an abnormal events that may constitute a major security issue which need to a quick action to avoid public danger. Hence, any transmission delay of the captured video can be crucial. Therefore, it is very important to ensure a non-delay transmission of the videos that present crucial information. However, in multimedia wireless sensors network, Transmission of multimedia content (i.e., video streams) over wireless sensors require transmission bandwidth that supports higher magnitude orders than those supported by currently available sensors. To this end, captured videos need to be compressed in order to minimize the data amount to be transmitted over the network. Lossy video coding offers a high compression rate that reduces significantly the video size. However, video coding may affect the video quality by removing some information from the original video, thereby reducing its size.

Based on these facts, we have dealt in this thesis with the global quality of the intelligent video surveillance system as a three parts problem which are video coding, the video quality assessment and the network quality of service.

6.1 Summary of contribution

As we deal with intelligent video surveillance systems, we have chosen to focus on new emerging kind of systems based on multiview cameras. More precisely, we focused on the stereoscopic coding issue as it is a particular case of the general multiview video surveillance system. Thus, as a first contribution in this thesis, a new stereoscopic view coding scheme that can be generalized to the multiview has been proposed. The second contribution of the thesis is the design of new quality-based intelligent video surveillance architecture where the video quality assessment is its key factor. This architecture is based on a smart video quality assessment box that can detect, classify and identify every distortion that may be present in the video. Therefore, a video surveillance oriented database with the most common distortions has been created in order to facilitate the implementation of the smart video quality assessment box.

The final contribution of this thesis is the implementation of efficient scheduling model based on priority of traffics for multimedia sensors that mitigates the waiting time while optimizing the throughput following the priority profile.

6.2 Perspectives and future work

Even through that this thesis proposes efficient approaches, still some areas and direction need to be improved and enhanced. In the following, we list some directions that can be investigated in future works:

- **Visual data coding** The proposed stereoscopic view coding scheme needs to be generalised to the context of multi-view coding. Moreover, other perceptual criteria could be investigated for the design of the different lifting operators.

- **VQA and Intelligent video surveillance system** The constructed database is an important step in facilitating development of new video quality assessment metric in the context of public visual surveillance. Moreover, a learning algorithm for distortions detection and identification for the smart quality assessment box in the proposed smart video surveillance architecture will be investigated in a future work.

- **Quality of Service in Wireless Multimedia Sensor Networks** These results open the door to future research in multimedia sensors network aiming to increase throughput, mitigate delay and optimize energy consumption by selecting channel and adapting bandwidth using intelligent mechanisms.

List of Publications

1. I. Bezzine, M. Kaaniche, Saadi Boudjit, Azeddine Beghdadi , « Sparse optimization of non separable vector lifting scheme for stereo image coding », J. Visual Communication and Image Representation 57: 283-293 (2018).

2. I. Bezzine, M. Kaaniche, N. Al-Maadeed, S. Boudjit, A. Beghdadi, « › Joint optimization of lifting operators for stereo image coding « , CORESA2018, 12-14 November 2018, Poitiers, France.

3. A. Manirabona, I. Bezzine, Saadi Boudjit, M. Kaaniche, Azeddine Beghdadi, « A Priority scheduling strategy for Retrial Queues in Wireless Multimedia Sensor Networks.» Wireless Networks Spring Journal (Submitted October 2019).

4. I. Bezzine, Z.A. Khan, N. Almaadeed, M. Kaaniche, S. Almaadeed, S. Boudjit, A. Bouridane, F.A. Cheikh, A. Beghdadi, « Video quality assessment dataset for smart public security systems» INMIC2020.

5. A. Beghdadi, I. Bezzine, M.A. Qureshi, M. Kaaniche, S. Boudjit « A Perceptual Quality-driven Video Surveillance System» INMIC2020.

Bibliography

[1] Valérie Gouaillier and A Fleurant. “Intelligent video surveillance: Promises and challenges”. In: *Technological and commercial intelligence report, CRIM and Technôpole Defence and Security* 456 (2009), p. 468 *Cited on pages 2, 16, 17.*

[2] E Wallace and C Diffley. “CCTV control room ergonomics”. In: *Published by Police Scientific Development Branch of the Home Office, Publication* 14/98 (1988) *Cited on page 2.*

[3] Gavin JD Smith. “Behind the screens: Examining constructions of deviance and informal practices among CCTV control room operators in the UK”. In: *Surveillance & Society* 2.2/3 (2004) *Cited on page 3.*

[4] Hannah M Dee and Sergio A Velastin. “How close are we to solving the problem of automated visual surveillance?” In: *Machine Vision and Applications* 19.5-6 (2008), pp. 329–343 *Cited on page 3.*

[5] Jianguo Chen, Kenli Li, Qingying Deng, Keqin Li, and S Yu Philip. “Distributed deep learning model for intelligent video surveillance systems with edge computing”. In: *IEEE Transactions on Industrial Informatics* (2019) *Cited on page 4.*

[6] Shaogang Gong, Chen Change Loy, and Tao Xiang. “Security and surveillance”. In: *Visual analysis of humans*. Springer, 2011, pp. 455–472 *Cited on page 13.*

[7] Robert T Collins, Alan J Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt, et al. “A system for video surveillance and monitoring”. In: *VSAM final report* 2000 (2000), pp. 1–68 *Cited on page 14.*

[8] Doug Washburn, Usman Sindhu, Stephanie Balaouras, Rachel A Dines, N Hayes, and Lauren E Nelson. “Helping CIOs understand “smart city” initiatives”. In: *Growth* 17.2 (2009), pp. 1–17 *Cited on page 14.*

[9] Ismail Haritaoglu, David Harwood, and Larry S. Davis. “W/sup 4: real-time surveillance of people and their activities”. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000), pp. 809–830 *Cited on page 14.*

[10] Omar Javed, Zeeshan Rasheed, Orkun Alatas, and Mubarak Shah. “KNIGHT/spl trade: a real time surveillance system for multiple and non-overlapping cameras”. In: *2003 International Conference on Multimedia and Expo. ICME’03. Proceedings (Cat. No. 03TH8698).* Vol. 1. IEEE. 2003, pp. I–649 *Cited on page 14.*

[11] Fereshteh Falah Chamasemani, Lilly Suriani Affendey, et al. “Systematic review and classification on video surveillance systems”. In: *International Journal of Information Technology and Computer Science (IJITCS)* 5.7 (2013), p. 87 *Cited on page 18.*

[12] Teddy Ko. “A survey on behavior analysis in video surveillance for homeland security applications”. In: *2008 37th IEEE Applied Imagery Pattern Recognition Workshop.* IEEE. 2008, pp. 1–8 *Cited on pages 18, 21, 24, 112.*

[13] Kosmas Dimitropoulos, Theodoros Semertzidis, and Nikolaos Grammalidis. “Video and signal based surveillance for airport applications”. In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance.* IEEE. 2009, pp. 170–175 *Cited on page 18.*

[14] Sehchan Oh, Sunghyuk Park, and Changmu Lee. “A platform surveillance monitoring system using image processing for passenger safety in railway

station". In: *2007 International Conference on Control, Automation and Systems*. IEEE. 2007, pp. 394–398 *Cited on page 18.*

[15] Ying-Wen Bai, Zi-Li Xie, and Zong-Han Li. "Design and implementation of a home embedded surveillance system with ultra-low alert power". In: *IEEE Transactions on Consumer Electronics* 57.1 (2011), pp. 153–159 *Cited on page 18.*

[16] Marijn JH Loomans, Cornelis J Koeleman, and Peter HN De With. "Low-complexity wavelet-based scalable image & video coding for home-use surveillance". In: *IEEE Transactions on Consumer Electronics* 57.2 (2011), pp. 507–515 *Cited on page 18.*

[17] Jia-Luen Chua, Yoong Choon Chang, and Wee Keong Lim. "A simple vision-based fall detection technique for indoor video surveillance". In: *Signal, Image and Video Processing* 9.3 (2015), pp. 623–633 *Cited on page 18.*

[18] Shengke Wang, Long Chen, Zixi Zhou, Xin Sun, and Junyu Dong. "Human fall detection in surveillance video based on PCANet". In: *Multimedia tools and applications* 75.19 (2016), pp. 11603–11613 *Cited on page 18.*

[19] Abderrahmane Ezzahout and Rachid Oulad Haj Thami. "Conception and development of a video surveillance system for detecting, tracking and profile analysis of a person". In: *2013 3rd international symposium ISKO-Maghreb*. IEEE. 2013, pp. 1–5 *Cited on page 18.*

[20] Peng Zhang, Yanning Zhang, Tony Thomas, and Sabu Emmanuel. "Moving people tracking with detection by latent semantic analysis for visual surveillance applications". In: *Multimedia tools and applications* 68.3 (2014), pp. 991–1021 *Cited on page 18.*

[21] Lei Wang and Fuhai Li. "The design of real-time monitoring system for enterprises in complex environments". In: *2012 International Conference on Systems and Informatics (ICSAI2012)*. IEEE. 2012, pp. 306–314 *Cited on pages 18, 26.*

[22] Pablo Negri. “Estimating the queue length at street intersections by using a movement feature space approach”. In: *IET Image Processing* 8.7 (2014), pp. 406–416 *Cited on page 18.*

[23] Le An, Mehran Kafai, and Bir Bhanu. “Face recognition in multi-camera surveillance videos using dynamic Bayesian network”. In: *2012 Sixth International Conference on Distributed Smart Cameras (ICDSC)*. IEEE. 2012, pp. 1–6 *Cited on page 18.*

[24] Le An, Bir Bhanu, and Songfan Yang. “Unified face representation for individual recognition in surveillance videos”. In: *Wide Area Surveillance*. Springer, 2014, pp. 123–136 *Cited on page 18.*

[25] Deng-Yuan Huang, Chao-Ho Chen, Wu-Chih Hu, Shu-Chung Yi, Yu-Feng Lin, et al. “Feature-based vehicle flow analysis and measurement for a real-time traffic surveillance system”. In: *Journal of Information Hiding and Multimedia Signal Processing* 3.3 (2012), pp. 279–294 *Cited on page 18.*

[26] Alberto Amato and Vincenzo Di Lecce. “Semantic classification of human behaviors in video surveillance systems”. In: *WSEAS Transactions on Computers* 10.10 (2011), pp. 343–352 *Cited on page 18.*

[27] Sanghyuk Park and Chang D Yoo. “Video scene analysis and irregular behavior detection for intelligent surveillance system”. In: *2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE. 2012, pp. 577–581 *Cited on page 18.*

[28] Hua-Tsung Chen, Li-Wu Tsai, Hui-Zhen Gu, Suh-Yin Lee, and Bao-Shuh P Lin. “Traffic congestion classification for nighttime surveillance videos”. In: *2012 IEEE International Conference on Multimedia and Expo Workshops*. IEEE. 2012, pp. 169–174 *Cited on page 18.*

[29] Xinfeng Bao, Solmaz Javanbakhti, Svitlana Zinger, Rob Wijnhoven, and Peter HN de With. “Context-based object-of-interest detection for a generic traffic surveillance analysis system”. In: *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE. 2014, pp. 136–141 *Cited on page 18.*

[30] Xiling Luo, Yanxiong Wu, Yan Huang, and Jun Zhang. “Vehicle flow detection in real-time airborne traffic surveillance system”. In: *Transactions of the Institute of Measurement and Control* 33.7 (2011), pp. 880–897 *Cited on page 18.*

[31] Max Krüger, Jürgen Ziegler, and Kathrin Heller. “A generic bayesian network for identification and assessment of objects in maritime surveillance”. In: *2012 15th International Conference on Information Fusion*. IEEE. 2012, pp. 2309–2316 *Cited on page 18.*

[32] Zygmunt L Szpak and Jules R Tapamo. “Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set”. In: *Expert systems with applications* 38.6 (2011), pp. 6669–6680 *Cited on page 18.*

[33] CL Lai, JC Yang, and YH Chen. “A real time video processing based surveillance system for early fire and flood detection”. In: *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*. IEEE. 2007, pp. 1–6 *Cited on page 19.*

[34] Kimin Yun, Hawook Jeong, Kwang Moo Yi, Soo Wan Kim, and Jin Young Choi. “Motion interaction field for accident detection in traffic surveillance video”. In: *2014 22nd International Conference on Pattern Recognition*. IEEE. 2014, pp. 3062–3067 *Cited on page 19.*

[35] Dalia Coppi, Simone Calderara, and Rita Cucchiara. “Iterative active querying for surveillance data retrieval in crime detection and forensics”. In: (2011) *Cited on page 19.*

[36] Le Lv, Dongbin Zhao, and Zhijiang Fan. “Cheating behavior detection based-on pictorial structure model”. In: *Proceedings of the 33rd Chinese Control Conference*. IEEE. 2014, pp. 7274–7279 *Cited on page 19.*

[37] Amal Ben Hamida, Mohamed Koubaa, Chokri Ben Amar, and Henri Nicolas. “Toward scalable application-oriented video surveillance systems”. In: *2014 Science and Information Conference*. IEEE. 2014, pp. 384–388 *Cited on page 19.*

[38] Jun Luo, Jinqiao Wang, Huazhong Xu, and Hanqing Lu. “A real-time people counting approach in indoor environment”. In: *International Conference on Multimedia Modeling*. Springer. 2015, pp. 214–223 *Cited on page 20.*

[39] Yaning Wang and Hong Zhang. “Pedestrian detection and counting based on ellipse fitting and object motion continuity for video data analysis”. In: *International Conference on Intelligent Computing*. Springer. 2015, pp. 378–387 *Cited on page 20.*

[40] Satarupa Mukherjee, BaidyaNath Saha, Iqbal Jamal, Richard Leclerc, and Nilanjan Ray. “Anovel framework for automatic passenger counting”. In: *2011 18th IEEE International Conference on Image Processing*. IEEE. 2011, pp. 2969–2972 *Cited on page 20.*

[41] Govind Salvi. “An automated nighttime vehicle counting and detection system for traffic surveillance”. In: *2014 International Conference on Computational Science and Computational Intelligence*. Vol. 1. IEEE. 2014, pp. 131–136 *Cited on page 20.*

[42] Imran Saleemi and Mubarak Shah. “Multiframe many–many point correspondence for vehicle tracking in high density wide area aerial videos”. In: *International journal of computer vision* 104.2 (2013), pp. 198–219 *Cited on page 20.*

[43] Sung Chun Lee and Ram Nevatia. “Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system”. In: *Machine vision and applications* 25.1 (2014), pp. 133–143 *Cited on page 20.*

[44] Elif Bektüzün, Yiğit S Küçüksöz, and M Elif Karslıgil. “Real time tracking and detection of unusual circumstances of elderly people with RGB-d camera”. In: *2013 21st Signal Processing and Communications Applications Conference (SIU)*. IEEE. 2013, pp. 1–5 *Cited on page 20.*

[45] Lucia Maddalena and Alfredo Petrosino. “Stopped object detection by learning foreground model in videos”. In: *IEEE transactions on neural networks and learning systems* 24.5 (2013), pp. 723–735 *Cited on page 20.*

[46] Julfa Tuty and Bailing Zhang. “Simultaneous object tracking and classification for traffic surveillance”. In: *Proceedings of International Conference on Computer Science and Information Technology*. Springer. 2014, pp. 749–755 *Cited on pages 20, 24.*

[47] Chundi Mu, Jianbin Xie, Wei Yan, Tong Liu, and Peiqin Li. “A fast recognition algorithm for suspicious behavior in high definition videos”. In: *Multimedia Systems* 22.3 (2016), pp. 275–285 *Cited on page 21.*

[48] Hashem Alayed, Fotos Frangoudes, and Clifford Neuman. “Behavioral-based cheating detection in online first person shooters using machine learning techniques”. In: *2013 IEEE Conference on Computational Inteligence in Games (CIG)*. IEEE. 2013, pp. 1–8 *Cited on page 21.*

[49] Rafael Martınez Tomás, Susana Arias Tapia, Antonio Fernández Caballero, Sylvie Ratté, Alexandra González Eras, Patricia Ludena González, et al. “Identification of loitering human behaviour in video surveillance environments”. In: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2015, pp. 516–525 *Cited on page 21.*

[50] Matthieu Brulin, Henri Nicolas, and Christophe Maillet. “Analyse d’un trafic routier dans un contexte de vidéo surveillance.” In: 2012 *Cited on page 21.*

[51] Insaf Setitra and Slimane Larabi. “Background subtraction algorithms with post-processing: a review”. In: *2014 22nd International Conference on Pattern Recognition*. IEEE. 2014, pp. 2436–2441 *Cited on page 23.*

[52] Alper Yilmaz, Omar Javed, and Mubarak Shah. “Object tracking: A survey”. In: *Acm computing surveys (CSUR)* 38.4 (2006), 13–es *Cited on pages 24, 113.*

[53] Thomas B Moeslund, Adrian Hilton, and Volker Krüger. “A survey of advances in vision-based human motion capture and analysis”. In: *Computer vision and image understanding* 104.2-3 (2006), pp. 90–126 *Cited on pages 24, 113.*

[54] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. "A survey on visual surveillance of object motion and behaviors". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.3 (2004), pp. 334–352 *Cited on pages 24, 26, 113.*

[55] Meng Li, Chuliang Wei, Ye Yuan, and Zemin Cai. "A survey of video object tracking". In: *International Journal of Control and Automation* 8.9 (2015), pp. 303–312 *Cited on pages 24, 113.*

[56] Rudra N Hota, Vijendran Venkoparao, and Anupama Rajagopal. "Shape based object classification for automated video surveillance with feature selection". In: *10th International Conference on Information Technology (ICIT 2007).* IEEE. 2007, pp. 97–99 *Cited on page 24.*

[57] Yao-Te Tsai, Huang-Chia Shih, and Chung-Lin Huang. "Multiple human objects tracking in crowded scenes". In: *18th International Conference on Pattern Recognition (ICPR'06).* Vol. 3. IEEE. 2006, pp. 51–54 *Cited on page 24.*

[58] Jae-Yeong Lee and Wonpil Yu. "Visual tracking by partition-based histogram backprojection and maximum support criteria". In: *2011 IEEE International Conference on Robotics and Biomimetics.* IEEE. 2011, pp. 2860–2865 *Cited on page 24.*

[59] Omar Javed and Mubarak Shah. "Tracking and object classification for automated surveillance". In: *European Conference on Computer Vision.* Springer. 2002, pp. 343–357 *Cited on page 25.*

[60] Himani S Parekh, Darshak G Thakore, and Udesang K Jaliya. "A survey on object detection and tracking methods". In: *International Journal of Innovative Research in Computer and Communication Engineering* 2.2 (2014), pp. 2970–2978 *Cited on page 25.*

[61] Hitesh A Patel and Darshak G Thakore. "Moving object tracking using kalman filter". In: *International Journal of Computer Science and Mobile Computing* 2.4 (2013), pp. 326–332 *Cited on page 25.*

[62] Swantje Johnsen and Ashley Tews. “Real-time object tracking and classification using a static camera”. In: *Proceedings of IEEE International Conference on Robotics and Automation, workshop on People Detection and Tracking*. Citeseer. 2009 *Cited on page 25.*

[63] Szabolcs Sergyán. “Color content-based image classification”. In: *5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics*. Citeseer. 2007, pp. 427–434 *Cited on page 25.*

[64] T Mahalingam and M Mahalakshmi. “Vision based moving object tracking through enhanced color image segmentation using Haar classifiers”. In: *Trendz in Information Sciences & Computing (TISC2010)*. IEEE. 2010, pp. 253–260 *Cited on page 25.*

[65] Navneet Dalal and Bill Triggs. “Histograms of oriented gradients for human detection”. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*. Vol. 1. IEEE. 2005, pp. 886–893 *Cited on page 25.*

[66] Azhar Mohd Ibrahim, AA Shafie, and MM Rashid. “Human identification system based on moment invariant features”. In: *2012 International Conference on Computer and Communication Engineering (ICCCE)*. IEEE. 2012, pp. 216–221 *Cited on page 26.*

[67] Yi-Ling Chen, Tse-Shih Chen, Tsiao-Wen Huang, Liang-Chun Yin, Shiou-Yaw Wang, and Tzi-cker Chiueh. “Intelligent urban video surveillance system for automatic vehicle detection and tracking in clouds”. In: *2013 IEEE 27th international conference on advanced information networking and applications (AINA)*. IEEE. 2013, pp. 814–821 *Cited on page 26.*

[68] Aaron F. Bobick and James W. Davis. “The recognition of human movement using temporal templates”. In: *IEEE Transactions on pattern analysis and machine intelligence* 23.3 (2001), pp. 257–267 *Cited on pages 26, 114.*

[69] Lee W Campbell and Aaron F Bobick. “Recognition of human body motion using phase space constraints”. In: *Proceedings of IEEE International Conference on Computer Vision*. IEEE. 1995, pp. 624–630 *Cited on pages 27, 114.*

[70] Yuntao Cui, Daniel L Swets, and John J Weng. “Learning-based hand sign recognition using SHOSLIF-M”. In: *Proceedings of IEEE International Conference on Computer Vision*. IEEE. 1995, pp. 631–636 *Cited on pages 27, 114.*

[71] Trevor Darrell and Alex Pentland. “Space-time gestures”. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 1993, pp. 335–340 *Cited on pages 27, 114.*

[72] Andrew David Wilson and Aaron F Bobick. “Learning visual behavior for gesture analysis”. In: *Proceedings of International Symposium on Computer Vision-ISCV*. IEEE. 1995, pp. 229–234 *Cited on pages 27, 114.*

[73] Jim Little and Jeffrey Boyd. “Describing motion for recognition”. In: *Proceedings of International Symposium on Computer Vision-ISCV*. IEEE. 1995, pp. 235–240 *Cited on pages 27, 115.*

[74] Irfan A. Essa and Alex Paul Pentland. “Coding, analysis, interpretation, and recognition of facial expressions”. In: *IEEE transactions on pattern analysis and machine intelligence* 19.7 (1997), pp. 757–763 *Cited on pages 27, 115.*

[75] I Bezzine, Mounir Kaaniche, Saadi Boudjit, and Azeddine Beghdadi. “Sparse optimization of non separable vector lifting scheme for stereo image coding”. In: *Journal of Visual Communication and Image Representation* 57 (2018), pp. 283–293 *Cited on page 30.*

[76] Poornima Krishnan and S Naveen. “RGB-D face recognition system verification using kinect and FRAV3D databases”. In: *Procedia Computer Science* 46 (2015), pp. 1653–1660 *Cited on page 32.*

[77] D. W. Hansen, H. S. Hansen, M. Kirschmeyer, R. Larsen, and D. Silvestre. “Cluster tracking with Time-of-Flight cameras”. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.* June 2008, pp. 1–6. DOI: 10.1109/CVPRW.2008.4563156 *Cited on page 32.*

[78] Davide Silvestre. “Video surveillance using a time-of-flight camera”. In: *Master’s thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU* (2007) *Cited on page 32.*

[79] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. “Human posture recognition with a time-of-flight 3D sensor for in-home applications”. In: *Expert Systems with Applications* 40.2 (2013), pp. 744–751 *Cited on page 32.*

[80] Rodrigo Ibanez, Alvaro Soria, Alfredo Teyseyre, and Marcelo Campo. “Easy gesture recognition for Kinect”. In: *Advances in Engineering Software* 76 (2014), pp. 171–180 *Cited on page 32.*

[81] JL Raheja, M Minhas, D Prashanth, T Shah, and A Chaudhary. “Robust gesture recognition using Kinect: A comparison between DTW and HMM”. In: *Optik* 126.11-12 (2015), pp. 1098–1104 *Cited on page 32.*

[82] Jake K Aggarwal and Lu Xia. “Human activity recognition from 3d data: A review”. In: *Pattern Recognition Letters* 48 (2014), pp. 70–80 *Cited on page 32.*

[83] Roanna Lun and Wenbing Zhao. “A survey of applications and human motion recognition with microsoft kinect”. In: *International Journal of Pattern Recognition and Artificial Intelligence* 29.05 (2015), p. 1555008 *Cited on page 32.*

[84] Can Wang and Hong Liu. “Unusual events detection based on multi-dictionary sparse representation using Kinect”. In: *2013 IEEE International Conference on Image Processing.* IEEE. 2013, pp. 2968–2972 *Cited on page 32.*

[85] Rami Alazrai, Yaser Mowafi, and CS George Lee. “Anatomical-plane-based representation for human–human interactions analysis”. In: *Pattern Recognition* 48.8 (2015), pp. 2346–2363 *Cited on page 32.*

[86] Caroline Rougier, Edouard Auvinet, Jacqueline Rousseau, Max Mignotte, and Jean Meunier. “Fall detection from depth map video sequences”. In: *International conference on smart homes and health telematics*. Springer. 2011, pp. 121–128 *Cited on page 32.*

[87] Young-Sook Lee and Wan-Young Chung. “Visual sensor based abnormal event detection with moving shadow removal in home healthcare applications”. In: *Sensors* 12.1 (2012), pp. 573–584 *Cited on page 32.*

[88] Georgios Mastorakis and Dimitrios Makris. “Fall detection system using Kinect’s infrared sensor”. In: *Journal of Real-Time Image Processing* 9.4 (2014), pp. 635–646 *Cited on page 32.*

[89] Amol Patwardhan and Gerald Knapp. “Aggressive actions and anger detection from multiple modalities using Kinect”. In: *arXiv preprint arXiv:1607.01076* (2016) *Cited on page 32.*

[90] Bernd Girod, Anne Aaron, Shantanu Rane, and DR Monedero. “Distributed video coding”. In: *Proceedings of IEEE Special Issue on Advances in Video Coding and Delivery*. Vol. 93. 1. Citeseer. 2003, pp. 1–12 *Cited on page 33.*

[91] I. Feldmann, W. Waizenegger, N. Atzpadin, and O. Schreer. “Real-Time depth estimation for immersive 3D videoconferencing”. In: *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*. Tampere, June 2010, pp. 1–4 *Cited on page 33.*

[92] B. Sdiri, M. Kaaniche, F. A. Cheikh, A. Beghdadi, and O. J. Elle. “Efficient enhancement of stereo endoscopic images based on joint wavelet decomposition and binocular combination”. In: *IEEE Transactions on Medical Imaging* (July 2018), 13 pages *Cited on page 33.*

[93] A. Kadaikar, G. Dauphin, and A. Mokraoui. “Sequential block-based disparity map estimation algorithm for stereoscopic image coding”. In: *Elsevier Signal Processing: Image Communication* 39.PA (Nov. 2015), pp. 159–172 *Cited on page 34.*

[94] D. Palaz, I. Tosic, and P. Frossard. “Sparse stereo image coding with learned dictionaries”. In: *IEEE International Conference on Image Processing*. Quebec, Canada, Sept. 2011, 4 pages *Cited on page 34.*

[95] G. Dauphin, M. Kaaniche, and A. Mokraoui. “Block dependent dictionary based disparity compensation for stereo image coding”. In: *IEEE International Conference on Image Processing*. Quebec, Canada, Sept. 2015, 5 pages *Cited on page 34.*

[96] L. F. Lucas, N. M. Rodrigues, C. L. Pagliari, E. A. Silva, and S. M. Faria. “Recurrent pattern matching based stereo image coding using linear predictors”. In: *Multidimensional Systems and Signal Processing* 28.4 (Oct. 2017), pp. 1393–1416 *Cited on page 34.*

[97] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh. “Optimized residual image for stereo image coding”. In: *European Workshop on Visual Information Processing*. Paris, France, Dec. 2014, pp. 1–6 *Cited on pages 34, 53, 54.*

[98] I. Tabus. “Patch-Based Conditional Context Coding of Stereo Disparity Images”. In: *IEEE Signal Processing Letters* 21.10 (Dec. 2014), pp. 1220–1224 *Cited on page 34.*

[99] O. Woo and A. Ortega. “Stereo image compression based on disparity field segmentation”. In: *SPIE Conference on Visual Communications and Image Processing*. Vol. 3024. San Jose, California, Feb. 1997, pp. 391–402 *Cited on page 35.*

[100] W. Hachicha, A. Beghdadi, and F. A. Cheikh. “1D directional DCT-based stereo residual compression”. In: *European Signal Processing Conference*. Marrakech, Morocco, Sept. 2013, 5 pages *Cited on page 35.*

[101] M. S. Moellenhoff and M. W. Maier. “Characteristics of disparity-compensated stereo image pair residuals”. In: *Signal Processing: Image Communications* 14 (1998), pp. 49–55 *Cited on page 35.*

[102] M. S. Moellenhoff and M. W. Maier. “Transform coding of stereo image residuals”. In: *IEEE Transactions on Image Processing* 7.6 (June 1998), pp. 804–812 *Cited on page 35.*

[103] N. V. Boulgouris and M. G. Strintzis. “A family of wavelet-based stereo image coders”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 12.10 (Oct. 2002), pp. 898–903 *Cited on page 35.*

[104] R. Darazi, A. Gouze, and B. Macq. “Adaptive lifting scheme-based method for joint coding 3D-stereo images with luminance correction and optimized prediction”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing.* Taipei, Apr. 2009, pp. 917–920 *Cited on page 35.*

[105] A. Maalouf and M.-C. Larabi. “Bandelet-based stereo image coding”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing.* Dallas, Texas, United States, Mar. 2010, pp. 698–701 *Cited on page 35.*

[106] E. Le Pennec and S. Mallat. “Sparse geometric image representations with bandelets”. In: *IEEE Transactions on Image Processing* 14.4 (Apr. 2005), pp. 423–438 *Cited on page 35.*

[107] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet. “Vector lifting schemes for stereo image coding”. In: *IEEE Transactions on Image Processing* 18.11 (Nov. 2009), pp. 2463–2475 *Cited on pages 35, 36, 39.*

[108] O. Dhifallah, M. Kaaniche, and A. Benazza-Benyahia. “Efficient joint multiscale decomposition for color stereo image coding”. In: *European Signal Processing Conference.* Lisbon, Portugal, Sept. 2014, 5 pages *Cited on page 35.*

[109] M. Kaaniche, B. Pesquet-Popescu, and J.-C. Pesquet. "ℓ_1-ADAPTED NON SEPARABLE VECTOR LIFTING SCHEMES FOR STEREO IMAGE". In: *European Signal Processing Conference*. Bucharest, Romania, Aug. 2012, 5 pages *Cited on pages 36, 37, 52.*

[110] V. Chappelier and C. Guillemot. "Oriented wavelet transform for image compression and denoising". In: *IEEE Transactions on Image Processing* 15.10 (Oct. 2006), pp. 2892–2903 *Cited on page 36.*

[111] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet. "Non separable lifting scheme with adaptive update step for still and stereo image coding". In: *Elsevier Signal Processing: Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations* 91.12 (Jan. 2011), pp. 2767–2782 *Cited on pages 36, 37, 42, 47, 52.*

[112] M. Kaaniche, B. Pesquet-Popescu, A. Benazza-Benyahia, and J.-C. Pesquet. "Adaptive lifting scheme with sparse criteria for image coding". In: *EURASIP Journal on Advances in Signal Processing: Special Issue on New Image and Video Representations Based on Sparsity* 2012, 22 pages (Jan. 2012) *Cited on pages 36, 42, 44, 46.*

[113] Y. Xing, M. Kaaniche, B. Pesquet-Popescu, and F. Dufaux. "Sparse based adaptive non separable vector lifting scheme for holograms compression". In: *International Conference on 3D Imaging*. Liège, Belgium, Dec. 2015, 8 pages *Cited on page 36.*

[114] W. Sweldens. "The lifting scheme: a new philosophy in biorthogonal wavelet construction". In: *Wavelet Applications in Signal and Image Processing III, SPIE*. San-Diego, CA, USA, Sept. 1995, pp. 68–79 *Cited on page 36.*

[115] D. Taubman and M. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Norwell, MA, USA: Kluwer Academic Publishers, 2001 *Cited on page 37.*

[116] O. N. Gerek and A. E. Cetin. "Adaptive polyphase subband decomposition structures for image compression". In: *IEEE Transactions on Image Processing* 9.10 (Oct. 2000), pp. 1649–1660 *Cited on page 42.*

[117] A. Gouze, M. Antonini, M. Barlaud, and B. Macq. “Design of signal-adapted multidimensional lifting schemes for lossy coding”. In: *IEEE Transactions on Image Processing* 13.12 (Dec. 2004), pp. 1589–1603 *Cited on page 42.*

[118] J. Chen, J. Hou, and L.-P. Chau. “Light Field Compression With Disparity-Guided Sparse Coding Based on Structural Key Views”. In: *IEEE Transactions on Image Processing* 27.1 (Jan. 2018), pp. 314–324 *Cited on page 42.*

[119] H. Gish and J. N. Pierce. “Asymptotically efficient quantizing”. In: *IEEE Transactions on Information Theory* 14.5 (1969), pp. 676–683 *Cited on page 42.*

[120] B. Usevitch. “Optimal bit allocation for biorthogonal wavelet coding”. In: *Data Compression Conference*. Snowbird, USA, Mar. 1996, pp. 387–395 *Cited on page 43.*

[121] S. Parrilli, M. Cagnazzo, and B. Pesquet-Popescu. “Distortion evaluation in transform domain for adaptive lifting schemes”. In: *International Workshop on Multimedia Signal Processing*. Cairns, Queensland, Australia, Oct. 2008, 6 pages *Cited on page 43.*

[122] J.-J. Moreau. “Proximité et dualité dans un espace hilbertien”. In: *Bulletin de la Societé Mathématique de France* 93 (1965), pp. 273–288 *Cited on page 44.*

[123] C. Chaux, P. Combettes, J.-C. Pesquet, and V. Wajs. “A variational formulation for framebased inverse problems”. In: *Inverse Problems* 23.4 (2007), pp. 1495–1518 *Cited on page 44.*

[124] P. L. Combettes and V. R. Wajs. “Signal Recovery by Proximal Forward-Backward Splitting”. In: *Multiscale Modeling and Simulation* 4.4 (2005), pp. 1168–1200 *Cited on page 44.*

[125] H. Hirschmüller and D. Scharstein. “Evaluation of cost functions for stereo matching”. In: *International Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN, USA, June 2007, 8 pages *Cited on page 52.*

[126] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling. “High-resolution stereo datasets with subpixel-accurate ground truth”. In: *German Conference on Pattern Recognition*. Münster, Germany, Sept. 2014, 12 pages *Cited on page 52.*

[127] G. Bjontegaard. *Calculation of average PSNR differences between RD curves*. Tech. rep. Austin, TX, USA: ITU SG16 VCEG-M33, Apr. 2001 *Cited on page 54.*

[128] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: From error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612 *Cited on page 55.*

[129] Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao. “A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications”. In: *IEEE Internet of Things Journal* 4 (2017), pp. 1125–1142. DOI: 10.1109/JIOT.2017.2683200\\ *Cited on page 62.*

[130] Pallavi Sethi and Smruti R. Sarangi. “Internet of Things: Architectures, Protocols, and Applications”. In: *Journal of Electrical and Computer Engineering* 2017 (2017), p. 25. DOI: https://doi.org/10.1155/2017/9324035 *Cited on page 62.*

[131] Litun Patra and Udai Pratap Rao. “Internet of Things Architecture, applications, security and other major challenges”. In: *3rd International Conference on Computing for Sustainable Global Development (INDIACom 2016)*. New Delhi, India, 16-18 March 2016 201 *Cited on page 62.*

[132] A. Manirabona, S. Boudjit, and L. C. Fourati. “A Priority-Weighted Round Robin scheduling strategy for a WBAN based healthcare monitoring system”. In: *13th IEEE Annual Consumer Communications and Networking Conference, CCNC 2016*. Las Vegas, NV, USA: IEEE Computer Society, Jan. 2016, pp. 224–229 *Cited on pages 62, 69.*

[133] Lutful Karim, Nidal Nasser, Tarik Taleb, and Abdullah Alqallaf. “An efficient priority packet scheduling algorithm for Wireless Sensor Network”. In: *2012 IEEE International Conference on Communications (ICC)*. Ottawa, ON, Canada, Oct. 2012 *Cited on page 62.*

[134] Octav Chipara, Chenyang Lu, and Gruia-Catalin Roman. “Real-time Query Scheduling for Wireless Sensor Networks”. In: *28th IEEE International Real-Time Systems Symposium (RTSS 2007)*. Tucson, AZ, USA, Mar. 2007 *Cited on page 62.*

[135] Islam T. Almalkawi, Manel Guerrero Zapata, Jamal N. Al-Karaki, and Julian Morillo-Pozo. “Wireless Multimedia Sensor Networks: Current Trends and Future Directions”. In: *Sensors 2010* 10 (2010), pp. 6662–6717 *Cited on page 62.*

[136] M. Alaei and J. M. Barcelo-Ordinas. “Priority-Based Node Selection and Scheduling for Wireless Multimedia Sensor Networks”. In: *2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*. Niagara Falls, NU, Canada, Nov. 2010. DOI: 10.1109/WIMOB.2010.5644981 *Cited on pages 62, 63.*

[137] Nasim Abbas, Fengqi Yu, and Yang Fan. “Intelligent Video Surveillance Platform for Wireless Multimedia Sensor Networks”. In: *Applied Sciences. 2018* 8.348 (2018). DOI: 10.3390/app8030348 *Cited on pages 62, 63.*

[138] Ahmed Salim, Walid Osamy, and Ahmed M. Khedr. “Effective Scheduling Strategy in Wireless Multimedia Sensor Networks for Critical Surveillance Applications”. In: *International Journal of Applied Mathematics and Information Sciences* 12.1 (2018), pp. 101–111. DOI: http://dx.doi.org/10.18576/amis/120109 *Cited on pages 62, 63.*

[139] Xue Wang, Sheng Wang, Junjie Ma, and Xinyao Sun. “Energy-aware Scheduling of Surveillance in Wireless Multimedia Sensor Networks”. In: *,Sensors 2010* 10 (2010), pp. 3100–3125. DOI: 10.3390/s100403100 *Cited on pages 62, 63.*

[140] Imane Horiya Brahmi, Soufiene Djahel, Damien Magoni, and John Murphy. "Messages Prioritization in IEEE 802.15.4 based WSNs for Roadside Infrastructure". In: *2014 International Conference on Connected Vehicles and Expo (ICCVE)*. Vienna, Austria, Mar. 2014. DOI: 10.1109/ICCVE.2014.7297521 *Cited on page 63.*

[141] Elham Karimi and Behzad Akbari. "Improving Video Delivery over Wireless Multimedia Sensor Networks Based on Queue Priority Scheduling". In: *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*. Wuhan, China, 23-25 Sept 2017. DOI: 10.1109/wicom.2011.6040552 *Cited on page 63.*

[142] Shu Fan. "A Cross-Layer Optimization QoS Scheme in Wireless Multimedia Sensor Networks". In: *Algorithms* 12.68 (2019). DOI: 10.3390/a12040068 *Cited on page 63.*

[143] Nidal Nasser, Lutful Karim, and Tarik Taleb. "Dynamic Multilevel Priority Packet Scheduling Scheme for Wireless Sensor Network". In: *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS* 12 (2013) *Cited on page 63.*

[144] IEEE Standard for Local and metropolitan area networks Part 15.4. *Low-Rate Wireless Personal Area Networks (LR-WPANs) 2006.* 2006 *Cited on pages 64, 66, 67.*

[145] Vladimir Hottmar and Bohumil Adamec. "Analytical Model of a Weighted Round Robin Service System". In: *Journal of Electrical and Computer Engineering* 2012 (2012), p. 6. DOI: doi.org/10.1155/2012/374961 *Cited on page 69.*

[146] Y. Zhang and P.O. Harrison. "Performance of a Priority-Weighted Round Robin Mechanism for Differentiated Service Networks". In: *Proceedings of 16th International Conference on Computer Communications and Networks.* Honolulu, 13-16 Aug. 2007 ICCCN 2007, pp. 1198–1203. DOI: 10.1109/ICCCN.2007.4317983 *Cited on page 69.*

[147] M.L. Pinedo. *Scheduling: theory, algorithms, and systems*. Springer, 2012 *Cited on page 69.*

[148] Tsung-Yu Tsai, Yao-Liang Chung, and Zsehong Tsai. "Communications and Networking". In: ed. by Jun Peng (Ed.) InTech, 2010. Chap. Introduction to Packet Scheduling Algorithms for Communication Networks. DOI: 10.5772/10167 *Cited on page 69.*

[149] William Stallings. *Operating Systems Internals and Design Principles*. New Jersey: Prentice Hall, 2012 *Cited on page 69.*

[150] Y. Zhang and P.G. Harrison. "Performance of a Priority-Weighted Round Robin Mechanism for Differentiated Service Networks". In: *16th International Conference on Computer Communications and Networks*. Honolulu, HI, USA, 13-16 Aug. 2007. DOI: 10.1109/ICCCN.2007.4317983 *Cited on page 70.*

[151] Ji-Young Kwak qnd Ji-Seung Nam and Doo-Hyun Kim. "A Modified Dynamic Weighted Round Robin Cell Scheduling Algorithm". In: *ETRI Journal* 24 (2002) *Cited on page 70.*

[152] K. Chen and K. Chen. "Performance Evaluation by Simulation and Analysis with Applications to Computer Networks". In: ed. by K. Chen (Ed.) 2015. Chap. The M/G/1 Queues. DOI: 10.1002/9781119006190.ch10 *Cited on page 71.*

[153] Moshe Zukerman. *Introduction to Queueing Theory and Stochastic Teletraffic Models*. Tech. rep. EE Department, City University of Hong Kong, Copyright M. Zukerman, 2000-2015 *Cited on pages 72, 73, 76.*

[154] Gautam Jain and Karl Sigman. "A Pollaczek-Khintchine Formula for M/G/1 Queues with Disasters". In: *Journal of Applied Probability* 33.4 (1996), pp. 1191–1200 *Cited on page 72.*

[155] Tetsuya Takine. "Distributional Form of Little's Law for FIFO Queues with Multiple Markovian Arrival Streams and Its Application to Queues

with Vacations". In: *Queueing Systems* 37.1-3 (Mar. 2001), pp. 31–63 *Cited on page 72.*

[156] Network Simulator 3. 2019 *Cited on page 77.*

[157] Riley G.F. and Henderson T.R. "Modeling and Tools for Network Simulation". In: ed. by Gross J. (eds) Wehrle K. Gunes M. Springer, Berlin, Heidelberg, 2010. Chap. The ns-3 Network Simulator. DOI: `https://doi.org/10.1007/978-3-642-12331-3-2` *Cited on page 77.*

[158] Amira Ben Mabrouk and Ezzeddine Zagrouba. "Abnormal behavior recognition for intelligent video surveillance systems: A review". In: *Expert Systems with Applications* 91 (2018), pp. 480–491 *Cited on page 85.*

[159] PLM Bouttefroy, A Bouzerdoum, SL Phung, and A Beghdadi. "Abnormal behavior detection using a multi-modal stochastic learning approach". In: *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE. 2008, pp. 121–126 *Cited on pages 85, 110.*

[160] Vlad Hosu, Franz Hahn, Mohsen Jenadeleh, Hanhe Lin, Hui Men, Tamas Sziranyi, Shujun Li, and Dietmar Saupe. *The Konstanz Natural Video Database*. 2017 *Cited on page 86.*

[161] Vlad Hosu, Franz Hahn, Mohsen Jenadeleh, Hanhe Lin, Hui Men, Tamas Sziranyi, Shujun Li, and Dietmar Saupe. "The Konstanz natural video database (KoNViD-1k)". In: *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. 2017, pp. 1–6 *Cited on page 86.*

[162] Francesca De Simone, Matteo Naccari, Marco Tagliasacchi, Frederic Dufaux, Stefano Tubaro, and Touradj Ebrahimi. "Subjective assessment of H. 264/AVC video sequences transmitted over a noisy channel". In: *2009 International Workshop on Quality of Multimedia Experience*. IEEE. 2009, pp. 204–209 *Cited on page 86.*

[163] Francesca De Simone, Marco Tagliasacchi, Matteo Naccari, Stefano Tubaro, and Touradj Ebrahimi. “A H. 264/AVC video database for the evaluation of quality metrics”. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2010, pp. 2430–2433 *Cited on page 86.*

[164] Chao Chen, Lark Kwon Choi, Gustavo de Veciana, Constantine Caramanis, Robert W Heath, and Alan C Bovik. “A dynamic system model of time-varying subjective quality of video streams over HTTP”. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2013, pp. 3602–3606 *Cited on page 86.*

[165] Chao Chen, Xiaoqing Zhu, Gustavo De Veciana, Alan C Bovik, and Robert W Heath. “Adaptive video transmission with subjective quality constraints”. In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, pp. 2477–2481 *Cited on page 86.*

[166] Cheng hsin Hsu and Mohamed Hefeeda. “Video quality for face detection, recognition, and tracking”. In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 7.1 (2011) *Cited on pages 87, 108.*

[167] A. Beghdadi, M. Asim, N. Almaadeed, and M. A. Qureshi. “Towards the design of smart video-surveillance system”. In: *2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*. Aug. 2018, pp. 162–167. DOI: 10.1109/AHS.2018.8541480 *Cited on pages 87, 108.*

[168] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack. “Study of subjective and objective quality assessment of video”. In: *IEEE transactions on Image Processing* 19.6 (2010), pp. 1427–1441 *Cited on page 90.*

[169] Noor Almaadeed, Muhammad Asim, Somaya Al-Maadeed, Ahmed Bouridane, and Azeddine Beghdadi. “Automatic detection and classification of audio events for road surveillance applications”. In: *Sensors* 18.6 (2018), p. 1858 *Cited on pages 91, 108.*

[170] WebCam. https://webcam.nl/media/ *Cited on page 93.*

[171] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana. "An information fidelity criterion for image quality assessment using natural scene statistics". In: *IEEE Transactions on image processing* 14.12 (2005), pp. 2117–2128 *Cited on page 97.*

[172] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain". In: *IEEE Transactions on image processing* 21.12 (2012), pp. 4695–4708 *Cited on page 97.*

[173] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. "Making a "completely blind" image quality analyzer". In: *IEEE Signal Processing Letters* 20.3 (2012), pp. 209–212 *Cited on page 97.*

[174] Hantao Liu and Ingrid Heynderickx. "Visual attention in objective image quality assessment: Based on eye-tracking data". In: *IEEE Transactions on Circuits and Systems for Video Technology* 21.7 (2011), pp. 971–982 *Cited on page 110.*

[175] Zhou Wang, Ligang Lu, and Alan C Bovik. "Video quality assessment based on structural distortion measurement". In: *Signal processing: Image communication* 19.2 (2004), pp. 121–132 *Cited on page 111.*

[176] Jie Yao, Yongqiang Xie, Jianming Tan, Zhongbo Li, Jin Qi, and Lanlan Gao. "No-reference video quality assessment using statistical features along temporal trajectory". In: *Procedia Engineering* 29 (2012), pp. 947–951 *Cited on page 111.*

[177] Yingjie Xia, Mingzhe Zhu, Qianping Gu, Luming Zhang, and Xuelong Li. "Toward solving the Steiner travelling salesman problem on urban road maps using the branch decomposition of graphs". In: *Information Sciences* 374 (2016), pp. 164–178 *Cited on page 111.*

[178] Jake K Aggarwal and Michael S Ryoo. "Human activity analysis: A review". In: *ACM Computing Surveys (CSUR)* 43.3 (2011), pp. 1–43 *Cited on page 113.*

[179] Philippe Loic Marie Bouttefroy, Azeddine Beghdadi, Abdesselam Bouzerdoum, and Son Lam Phung. “Markov random fields for abnormal behavior detection on highways”. In: *2010 2nd European Workshop on Visual Information Processing (EUVIP)*. IEEE. 2010, pp. 149–154 *Cited on page 114.*

[180] Muhammad Atif Tahir, Fei Yan, Peter Koniusz, Muhammad Awais, Mark Barnard, Krystian Mikolajczyk, Ahmed Bouridane, and Josef Kittler. “A robust and scalable visual category and action recognition system using kernel discriminant analysis with spectral regression”. In: *IEEE Transactions on Multimedia* 15.7 (2013), pp. 1653–1664 *Cited on page 114.*

[181] Dima Damen and David Hogg. “Detecting carried objects from sequences of walking pedestrians”. In: *IEEE transactions on pattern analysis and machine intelligence* 34.6 (2011), pp. 1056–1067 *Cited on page 115.*

[182] Oluwatoyin P Popoola and Kejun Wang. “Video-based abnormal human behavior recognition—A review”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.6 (2012), pp. 865–878 *Cited on page 115.*

[183] Yong Shean Chong and Yong Haur Tay. “Abnormal event detection in videos using spatiotemporal autoencoder”. In: *International Symposium on Neural Networks*. Springer. 2017, pp. 189–196 *Cited on page 116.*

[184] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. “Abnormal event detection in videos using generative adversarial nets”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 1577–1581 *Cited on page 116.*

[185] Zeyu Ding, Yuxin Wang, Guanhong Wang, Danfeng Zhang, and Daniel Kifer. “Detecting violations of differential privacy”. In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 2018, pp. 475–489 *Cited on page 116.*

[186] Khiang Zarchi Htun and Sai Maung Maung Zaw. “Human identification system based on statistical gait features”. In: *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*. IEEE. 2018, pp. 508–512 *Cited on page 116.*

[187] Shruti Jadon and Mahmood Jasim. “Video Summarization using Keyframe Extraction and Video Skimming”. In: *arXiv preprint arXiv:1910.04792* (2019) *Cited on page 117.*

[188] Debi Prosad Dogra, Arif Ahmed, and Harish Bhaskar. “Smart video summarization using mealy machine-based trajectory modelling for surveillance applications”. In: *Multimedia Tools and Applications* 75.11 (2016), pp. 6373–6401 *Cited on page 117.*