

**ALMA MATER STUDIORUM  
UNIVERSITÀ DEGLI STUDI DI BOLOGNA**

---

**Dottorato di Ricerca in Ingegneria Elettronica, delle  
Telecomunicazioni e Tecnologie dell'Informazione**

Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione  
"Guglielmo Marconi"

**Ciclo XXVIII**

**Settore concorsuale di afferenza: 09/F2 - TELECOMUNICAZIONI**

**Settore scientifico disciplinare: ING-INF/03 - TELECOMUNICAZIONI**

**DATA PROCESSING AND FUSION  
FOR MULTI-SOURCE WIRELESS  
SYSTEMS**

*Presentata da:*  
SIMONE MORETTI

*Coordinatore Dottorato:*  
Prof. Ing.  
ALESSANDRO VANELLI  
CORALLI

*Relatore:*  
Chiar.mo Prof. Ing.  
MARCO CHIANI  
Prof. Ing.  
ENRICO PAOLINI

---

ESAME FINALE 2016

## **INDEX TERMS**

Camera Selection

Video Adaptation

Resource Allocation

LTE

QoE

*“Never say never.  
Because limits, like fears, are often just an illusion.”*  
Michael Jordan.



## Acronyms

**2D** Two-Dimensional

**3D** Three-Dimensional

**3G** Third Generation

**3GPP** 3<sup>rd</sup> Generation Partnership Project

**4G** Fourth Generation

**AoI** Area of Interest

**ALRE** Adaptive Low-rate Encoding

**AMC** Adaptive Modulation and Coding

**ARP** Allocation and Retention Priority

**AVC** Advanced Video Coding

**AWGN** Additive White Gaussian Noise

**BER** Bit Error Rate

**BL** Base Layer

**BLER** Block Error Rate

**BPSK** Binary Phase Shift Keying

**BS** Base Station

**CC** Coordination Center

**CGS** Coarse Grain Scalability

**CONCERTO** Content and cOntext aware delivery for iNteraCtive multimedia  
healthcaRe applications

**CP** Cyclic Prefix

**CQI** Class Quality Indicators

**CRA** Camera Ranking Algorithm

**CRDS** Camera Ranking for Dynamic Scenes

**CRSS** Camera Ranking for Static Scenes  
**CSI** Channel State Information  
**DFT** Discrete Fourier Transform  
**DL** Data Link  
**EA** Emergency Area  
**ECG** Electrocardiogram  
**eHealth** Electronic Health  
**eNodeB** Evolved Node B  
**EPC** Evolved Packet Core  
**ER** Equal rate  
**FAST** Focused Assessment with Sonography for Trauma  
**FEC** Forward Error Correction  
**FDD** Frequency Division Duplexing  
**FDMA** Frequency Division Multiple Access  
**FGS** Fine Granularity Scalability  
**FoV** Field of View  
**FP7** Seventh Framework Programme  
**fps** frame-per-second  
**FTP** File Transfer Protocol  
**GBR** Guaranteed Bit-rate  
**GOP** Groups of Pictures  
**GPS** Global Positioning System  
**ICT** Information and Communication Technologies  
**IDE** Interactive Development Environment

**IDR** Instantaneous Decoder Refresh

**I-FDMA** Interleaved-FDMA

**IoT** Internet of Things

**IP** Internet Protocol

**IPv6** Internet Protocol Version 6

**ISI** Inter-symbol Interference

**ISO/OSI** Open Systems Interconnection

**ITU** International Telecommunications Union

**JSVM** Joint Scalable Video Model

**KPI** Key Performance Index

**KTD** Kendall's Tau Distance

**LDPC** Low-density Parity-check

**L-FDMA** localized-FDMA

**LLR** Log-Likelihood Ratio

**LRE** Low-rate Encoding

**LTE** Long Term Evolution

**LTE-A** Long Term Evolution-Advanced

**M2M** Machine-to-Machine

**MAC** Media Access Control

**MGS** Medium Grain Scalability

**mHealth** Mobile Health

**MIMO** Multiple-input Multiple-output

**MMSE** Minimum Mean Square Error

**MOS** Mean Opinion Score

**MRC** Maximum Ratio Combining

**MSE** Mean Square Error

**MSM** Multi-source Management

**MTC** Machine Type Communication

**MWQ** Maximum Weighted Quality

**OFDM** Orthogonal Frequency Division Multiplexing

**OFDMA** Orthogonal Frequency Division Multiple Access

**PAPR** Peak-to-average Power Ratio

**PELR** Packet Error Loss Rate

**PEQ** Prioritized Equal Quality

**PCC** Pearson Correlation Coefficient

**PDB** Packet Delay Budget

**PF** Proportional Fair

**PHY** Physical

**PL-FEC** Packet-level Forward Error Correction

**PMOS** Pairwise Mean Opinion Score

**PRB** Physical Resource Block

**PSK** Phase Shift Keying

**PSNR** Peak Signal-to-noise Ratio

**QAM** Quadrature Amplitude Modulation

**QCI** QoS Class Identifier

**QF** Quality fair

**QoE** Quality of Experience

**QoS** Quality of Service

**QPSK** Quadrature Phase Shift Keying  
**QP** Quantization Parameter  
**Q-R** Quality-rate  
**R-D** Rate-distortion  
**RRA** Radio Resource Allocation  
**RRM** Radio Resource Management  
**RTLS** Real-Time Location System  
**RTP** Real-Time Protocol  
**SCC** Spearman's Correlation Coefficient  
**SC-FDMA** Single Carrier-Frequency Division Multiple Access  
**SNR** Signal-to-noise Ratio  
**SR** Source Ranking  
**SSIM** Structural Similarity  
**SVC** Scalable Video Coding  
**TCP** Transport Control Protocol  
**TDD** Time Division Duplexing  
**UBF** Uncorrelated Block Fading  
**UC** Use Case  
**UE** User Equipment  
**VSN** Visual Sensor Network  
**WDS** Wireless Distribution System  
**WKTD** Weighted Kendall's Tau Distance  
**WLAN** Wireless Local Area Network



# Contents

<b>Abstract</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>1 Transmission Of Heterogeneous Health-Related Multimedia Information From Emergency Areas</b>	<b>7</b>
1.1 Use Case 1: Ambulance And Emergency Area . . . . .	8
1.1.1 Expected Innovations In The Ambulance (Application Layer)	9
1.1.2 Expected Innovations At The eNodeB (DL and MAC Layers)	10
1.1.3 Expected Innovations At The eNodeB (PHY Layer) . . . . .	11
1.2 Use Case 2: Emergency Area With Multiple Casualties . . . . .	13
1.2.1 Expected Innovation At The Coordination Center (Application Layer) . . . . .	15
1.2.2 Expected Innovation At The eNodeB (DL And MAC Layers)	15
1.2.3 Expected Innovation In The Ambulance (Application Layer)	16
<b>2 Camera Selection Techniques For Large Camera Networks</b>	<b>21</b>
2.1 Introduction And State-of-the-art . . . . .	21
2.1.1 Video Transmissions In Multi-Camera Systems . . . . .	22
2.1.2 Embedded Processing . . . . .	22
2.1.3 Sensor Collaboration . . . . .	23
2.1.4 Sensor Management And Camera Sselection . . . . .	23
2.1.5 A Preliminary Camera Selection Technique In Emergency Areas . . . . .	24
2.2 Camera Ranking Techniques . . . . .	25
2.2.1 Scenario Description . . . . .	26
2.2.2 Camera Ranking for Static Scenes (CRSS) . . . . .	27
2.2.3 Camera Ranking for Dynamic Scenes (CRDS) . . . . .	29
2.2.4 Preliminary Tests Over The CRSS For Emergency Areas And Medical Contexts . . . . .	31

2.2.5	Experimental results . . . . .	37
2.2.6	Subjective Video Quality Assessment Results . . . . .	39
2.2.7	Camera Ranking Techniques Comparison . . . . .	40
2.2.8	Camera Ranking Validation . . . . .	41
2.2.9	Camera Ranking Evaluation Tools . . . . .	43
2.3	Conclusions . . . . .	49
<b>3</b>	<b>Video Adaptation And Transmission Techniques For mHealth</b>	<b>53</b>
3.1	Introduction And State-of-the-art . . . . .	53
3.1.1	Radio Resource Allocation And Scheduling In LTE Uplink . . . . .	55
3.1.2	Cross-layer Optimization For Video Streams . . . . .	58
3.2	Cross-layer Optimization For mHealth SVC Multiple Video Trans- mission over LTE Uplink . . . . .	61
3.2.1	System Architecture . . . . .	61
3.2.2	Video Coding And Adaptation . . . . .	63
3.2.3	Adaptation Algorithms . . . . .	64
3.2.4	LTE Uplink Model . . . . .	65
3.2.5	Numerical Results . . . . .	68
3.3	Content/Context-Aware Multiple Camera Selection and Video Adap- tation For mHealth . . . . .	70
3.3.1	System Architecture . . . . .	71
3.3.2	Multi Camera Ranking . . . . .	72
3.3.3	Video Coding And Adaptation . . . . .	74
3.3.4	Numerical Results . . . . .	76
3.4	An Optimization Framework for Content/Context-aware SVC Multiple Video Delivery In mHealth Emergency Applications . . . . .	80
3.4.1	System Architecture . . . . .	81
3.4.2	Multi Camera Ranking . . . . .	83
3.4.3	Joint Video Selection And SVC Adaptation . . . . .	85
3.4.4	Numerical Results . . . . .	94
3.5	Conclusions . . . . .	101
<b>4</b>	<b>Design And Implementation Of The System Simulator</b>	<b>103</b>
4.1	Simulator Architecture . . . . .	105
4.2	Macro Module Description . . . . .	106
4.2.1	Hospital 2 . . . . .	108
4.2.2	Emergency Area . . . . .	108
4.2.3	Servers And Clients . . . . .	109
4.3	Simulator Module Design . . . . .	110
4.3.1	Application Controller . . . . .	110

4.3.2	Multi-Source Management Module (Coordination Center) Module Description . . . . .	112
4.3.3	Multi-Source Management Module (Ambulance) . . . . .	113
4.3.4	Fourth Generation (4G) Radio Resource Management (RRM) Module . . . . .	114
4.3.5	Physical (PHY) - MIMO Modem And Frame De-Assembler Modules . . . . .	115
4.3.6	Radio Channel Module . . . . .	117
4.4	Simulation Results . . . . .	119
4.4.1	KPI Evaluated Through Simulations . . . . .	119
4.4.2	Simulated Scenarios . . . . .	119
4.5	Conclusions . . . . .	158

<b>Conclusions</b>	<b>159</b>
--------------------	------------



# Abstract

The constant evolution of the telecommunication technologies is one fundamental aspect that characterizes the modern era. In the context of healthcare and security, different scenarios are characterized by the presence of multiple sources of information that can support a large number of innovative services. For example, in emergency scenarios, reliable transmission of heterogeneous information (health conditions, ambient and diagnostic videos) can be a valid support for managing the first-aid operations. The presence of multiple sources of information requires a careful communication management, especially in case of limited transmission resource availability. The objective of my Ph.D. activity is to develop new optimization techniques for multimedia communications, considering emergency scenarios characterized by wireless connectivity. Different criteria are defined in order to prioritize the available heterogeneous information before transmission. The proposed solutions are based on the modern concept of content/context awareness: the transmission parameters are optimized taking into account the informative content of the data and the general context in which the information sources are located. To this purpose, novel cross-layer adaptation strategies are proposed for multiple SVC videos delivered over wireless channel. The objective is to optimize the resource allocation dynamically adjusting the overall transmitted throughput to meet the actual available bandwidth. After introducing a realistic camera network, some numerical results obtained with the proposed techniques are showed. In addition, through numerical simulations the benefits are showed, in terms of QoE, introduced by the proposed adaptive aggregation and transmission strategies applied in the context of emergency scenarios. The proposed solution is fully integrated in European research activities, including the FP7 ICT project CONCERTO<sup>1</sup>. To implement, validate and demonstrate the functionalities of the proposed solutions, extensive transmission simulation campaigns are performed. Hence, the presented solutions are integrated on a common system simulator which is been developed within the CONCERTO project.

---

<sup>1</sup>FP7 ICT CONCERTO project, Grant no. 288502.



# Introduction

During the last century, significant scientific progress has more than doubled the average lifetime of developed countries inhabitants. This amazing result is the indirect consequence of X-rays discovery in 1901, of antibiotics in the late 20s, of advances in surgery with the first heart transplant in the late 60s, to name just a few representative examples. In the new millennium, the use of Information and Communication Technologies (ICT) for health will likely constitute the next disruptive big jump: this goal will be fostered by spreading eHealth services by addressing the delivery of interactive multimedia applications over wireless networks, which will be a fundamental component of next generation healthcare services. Tele-medicine applications play an increasingly important role in healthcare and provide tools that are indispensable for home healthcare, remote patient monitoring and disease management. ICT has reached a degree of maturity allowing us to envisage setting up a virtual hospital in a patient's home, relying near-real-time tele-consulting and diagnosis from a remote location and transmitting clinical data and multimedia medical content from one location to a large number of geographically dispersed locations. The proposed solutions are capable of speeding up diagnosis as well as therapeutic care delivery, while allowing healthcare providers to receive continuous assistance from specialised centres. At a technical level, interactive and real-time multimedia delivery and adaptation, a hot topic today, becomes even more challenging when considering remote healthcare applications, requiring high received quality for a correct diagnosis: flawless access to Two-Dimensional (2D) and Three-Dimensional (3D) images and videos while roaming has to be guaranteed for doctors and paramedics. These demanding specifications are further aggravated by the potential world-wide roaming of both the patients as well as of the highly-trained medical consultants. To address these challenges and to provide a high Quality of Experience (QoE) for medical personnel using these applications, new techniques for media content fusion, adaptation, delivery and interactive access have to be designed and validated. In its turn, medical images and videos require improved compression schemes and high performance protection techniques to be delivered in a reliable way. In this context, near-instantaneous adaptation is mandatory for coping with variable bandwidth

availability, error-prone links, etc., that may affect the received quality.

To bridge this gap, we design and validate new techniques for media content fusion, delivery and interactive access, while specifically addressing the strict requirements of healthcare services. The principal aim of my work is to provide high QoE for medics, which is a necessary condition for providing flawless medical diagnosing of the highest reliability. To this purpose, a system architecture is envisaged with the aim at satisfying the requirements of healthcare services and to deploy different multi-user scenarios considered in the real world. It is worth noting that the most part of the results and the proposed solutions are fully integrated in European research activities, in particular within the framework of the Seventh Framework Programme (FP7) ICT project Content and cOntext aware delivery for iNteraCtive multimedia healthcaRe applications (CONCERTO). The system architecture is described in Chapter 1 in which the principal constituent blocks are highlighted. In this blocks, optimization algorithms are integrated with the aim to:

- I provide an efficient multi-user selection and scheduling scheme based on specific medical quality requirements,
- II provide near real-time content and context-aware selection and source encoding adaptation strategies,
- III provide optimal Long Term Evolution (LTE)-based radio resource allocation framework according to the network conditions.

One of the most important problem arising in the emergency communication scenarios presented in Chapter 1 is related to the huge amount of multimedia information which should be transmitted and managed in conjunction with medical data. Considering, as an example, videos from an emergency area, it may be feasible to fuse, combine and adapt also the signals of numerous camera-phones thus generated in a spontaneous and distributed manner. The generated stream contains redundant information which could be exploited to increase the probability of a lossless video transmission and, on the other side, this information could be fused and/or selected for delivery in case of limited resources. For example, most of modern mobile phones include Global Positioning System (GPS) positioning units, gyroscopes and electronic compasses which can provide geographical and viewpoint information about pictures and videos being transmitted. A proper classification of the collected media information through data fusion techniques may allow remote users to interactively move through the different available cameras acquiring a more precise idea of a given crisis scenario.

Chapter 2 is dedicated to exploit camera ranking algorithms to logically organize and select multiple multimedia streams in the context of large camera

networks. In this regard, I conceive new camera ranking strategies for selecting and prioritizing multiple information sources constituted by multiple cameras with different positions and orientations. Spontaneous provision of source positioning information (*e.g.*, obtained through GPS receivers and electronic compasses) may be exploited in this process. This facilitates flexible end-user interactivity such as content navigation as well as viewpoint selection. To evaluate the effectiveness of the camera selection techniques, different sets of simulations are performed using cameras that are characterized by different spatial and temporal resolution. At the end of each simulation, the proposed camera ranking algorithms are validated through objective metrics (Peak Signal-to-noise Ratio (PSNR) and Structural Similarity (SSIM)) and subjective tests, such as Pairwise Mean Opinion Score (PMOS).

The long-distance transmission of health-related information (*e.g.*, from an ambulance to a remote hospital) is a challenging task, due to the variability and the limitations of the mobile radio link. In particular, the transmission of multiple video streams can improve the efficacy of the tele-consultation service, but requires a large bandwidth to meet the desired quality, not always guaranteed by the mobile network. Consequently, the optimization of the wireless network is necessary for transporting the large amounts of both multimedia and medical information, either in the downlink and in the uplink. In Chapter 3, I describe novel adaptation schemes that are designed for centralized uplink transmissions. In particular, I propose novel cross-layer adaptation strategies for multiple Scalable Video Coding (SVC) videos delivered over a single LTE channel, which dynamically adjusts the overall transmitted throughput to meet the actual available bandwidth. Moreover, specific constraints to interactive healthcare and safety applications like delay and reliability are introduced, which are neglected by existing approaches mostly trying to adapt the encoding rate to the available bandwidth. I intend to work on dedicated video coding algorithms, at a quality level that ensures proper diagnosis and guaranteeing secure transmission. The proposed adaptation strategies jointly interact with the camera ranking techniques I present in Chapter 2. After having introduced a realistic uplink scenario, I take into account media content-aware wireless delivery over LTE/Long Term Evolution-Advanced (LTE-A) networks. The particular objectives include the optimization of LTE uplink transmissions for video delivery, considering all relevant variables (*e.g.*, distance from base station, power adaptation, traffic priority, etc.) and decision techniques (*e.g.*, scheduling and resource allocation). Modern Single Carrier-Frequency Division Multiple Access (SC-FDMA) schemes are considered in this Chapter. Moreover, Media Access Control (MAC) level scheduling solutions for the uplink are investigated, capable to differently prioritise the data traffic in a shared wireless network and enabling interactive healthcare applications.

To implement, validate and demonstrate the functionalities of the proposed solutions, extensive transmission simulation campaigns are performed. For this purpose, Chapter 4 is dedicated on the integration of the presented solutions on a common system simulator which is been developed within the CONCERTO project. The development of a simulation platform turns out to be an excellent solution for validating and also stimulating the achievement of novel results. In particular, I focus on the design of the simulation architecture, as well as on the description of the main modules that we have included in the software framework to implement the camera ranking algorithms, the SVC adaptation techniques and the multi-user resource allocation and scheduling schemes for LTE uplinks proposed in Chapter 3.

# Chapter 1

## Transmission Of Heterogeneous Health-Related Multimedia Information From Emergency Areas

Tele-medicine applications play an increasingly important role in healthcare and provide tools that are indispensable for home healthcare, remote patient monitoring and disease management. ICT has reached a degree of maturity allowing us to envisage setting up a virtual hospital in a patient's home, relying near-real-time tele-consulting and diagnosis from a remote location and transmitting clinical data and multimedia medical content from one location to a large number of geographically dispersed locations. Advances in technology including wireless connectivity and mobile devices will give practitioners, medical centres and hospitals important new tools for managing patient care, electronic records and for the billing of medical services, in order to ultimately enable patients to have more control over their own well-being. While multimedia applications, such as for example 2D/3D imaging of anatomy, have become a routine part of doctors' everyday work, by contrast, remotely participating in life-saving operations by renowned experts remains an elusive future application-perhaps apart from a few fixed-network-based hospital-to-hospital communications. In this field, ICT acts a principal role for supporting novel solutions for speeding up diagnosis as well as therapeutic care delivery, while allowing healthcare providers to receive continuous assistance from specialised centres. In this regard, the multimedia transmission content from emergency zones to remote hospital is one of the most challenging scenarios. Two different use cases have been considered that are also been object of study within the CONCERTO project [7].

## 1.1 Use Case 1: Ambulance And Emergency Area

This use case considers the call of the emergency services and the transmission of image and video data to/from an ambulance in an emergency situation. Different types of image/video data have to be transmitted together with other vital medical signals such as Electrocardiogram (ECG) data, including live ultrasound image data streams/clips, while supporting a live video feed to assess the general condition of the patient, on the move. The availability of this information allows medics to decide on the choice of the most suitable hospital/department which should treat the patient, to prepare the hospitalisation, and/or to guide medical or paramedic staff during the operation to be performed on the patient. This scenario would involve the employment of both high-quality lossy as well as lossless compression methodologies and presents challenging requirements in terms of the overall reproduction quality and delay, according to transmission resources available. Example of video and images that could be sent from the ambulance to the hospital are:

- ECG and other vital signal,
- Ultrasound (images and video): if the ambulance is equipped with a portable ultrasound device, an ultrasonography can be performed on the patient on the way to the hospital, in order to check possible internal injuries or heart problems, and the acquired video sequence can be sent to the hospital via a mobile broadband platform. In particular, LTE/LTE-A is considered,
- High-definition live video from the ambulance or from the emergency as additional information for the hospital staff (doctors, nurses etc.). Visualising the patient during the connection with the hospital can in fact be useful for a better assessment of its status. Furthermore, in case emergency treatments or examinations are needed while the patient is in the ambulance, the high quality video link can be used to enable the specialists in the hospital to guide the treatment/examination (remotely guided ultrasonography, etc.).

The concept of the Use Case (UC)-1 is sketched in the Fig. 1.1 in which two main areas are identified, namely the *Emergency Area* and the *Hospital*. The data are directly generated on-board of the ambulance during the way to reach the hospital. Videos and images can be originated by multiple devices with different characteristics and can require different kind of compression: data can be either generated by ambient videos and medical one (*i.e.*, Ultrasound). At the hospital, instead, the nodes are mainly receivers and typically correspond to doctors that can receive data from the emergency area both through wireless networks of the hospital. To avoid overcrowding the figure, only user data transmissions are drawn,

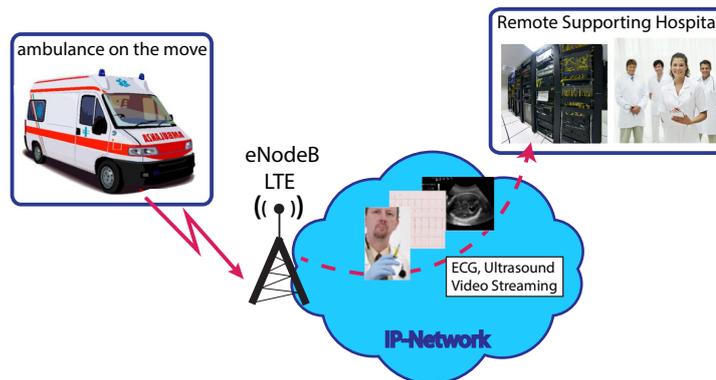


Figure 1.1: Ambulance and emergency area use case.

but several fluxes of cross layer data are also present and allow the control of the different transmissions [8].

### 1.1.1 Expected Innovations In The Ambulance (Application Layer)

The multimedia content will be adapted to the transmission conditions for delivering medical data in a reliable way and for performing efficient uplink video streaming satisfying the stringent delay requirements. In this terms, it is assumed the ambulance is equipped with a SVC adaptation module and it is capable to connect to the actual LTE cell. From a medical point of view, there is the need to know the state of health of the patient and therefore to collect all the related information. Regarding image/video interpretation, important aspects are: fast acquisition with real-time imaging, good quality of images, high resolution, accurate qualitative analysis and good interface characterisation. According to the above mentioned requirements, novel encoder/decoder schemes at the application layer have been developed. These blocks will be envisaged following the H.264/SVC standard. To this purpose, the processing unit on-board the ambulance is responsible to collect and aggregate the multimedia streams from ambient cameras and from medical devices available in the ambulance and to transmit them through the LTE link, toward the hospital. To perform its processing, the ambulance exploits the achievable throughput across the LTE link. The target video coding rates of the selected sources are jointly adapted to guarantee the optimal level of quality (in terms of PSNR and SSIM) compatible with the available radio resources and communicated to the application controller of each encoder. Again, novel radio resource allocation and multi-user scheduling schemes are provided.

### 1.1.2 Expected Innovations At The Evolved Node B (eNodeB) (Data Link (DL) And MAC Layers)

The DL layer is located between the physical and network layers and act as a mediator for data delivery between these two layers. In practice, the data link layer provides the functional and procedural means to transfer data between network entities and might provide the means to detect and possibly correct errors that occur in the physical layer. The main functionalities of the MAC layer include providing and controlling the medium access for devices, creation of the transmission frame, and the delivery of the transmission frame to base band entities. The data link layer plays an important role especially in the radio access network. With LTE, the link layer is in charge of providing the Quality of Service (QoS) mechanism, managing the user mobility and the radio resources. Optimizing the link layer or performing cross layer optimization across the link layer are expected to considerably improve the support of medical applications. Specific controlling functionalities at the DL Layer of LTE/LTE-A eNodeB are included, in charge of managing multi-user scheduling for uplink, allocating the most appropriate radio resources and adaptively selecting the physical transmission schemes (modulation and coding) capable to guarantee the required QoS. Healthcare related video traffic can be prioritized over other less critical traffic with different MAC layer QoS-solutions supported by most wireless standards. In particular, LTE provides a standardized traffic class differentiation based on sets of predefined QoS requirements. Each class is intended to provide a desired level of service for the selected traffic across the 3<sup>rd</sup> Generation Partnership Project (3GPP) network. Basically, the service classes are defined in terms of throughput, delay (*e.g.*, latency and jitter) and packet loss rate. In LTE the data traffic generated by each device is classified into bearers. A bearer is the basic traffic element that enables differentiated treatment according to specific QoS requirements. Each traffic class is identified by a scalar value, called QoS Class Identifier (QCI). Basically there are two major LTE traffic categories: guaranteed bit rate Guaranteed Bit-rate (GBR) and non-GBR. GBR are typically used for real-time services, such as conversational voice and video that have low latency and jitter tolerance. Non-GBR bearers do not need prefixed network bandwidth allocation. Non-GBR bearers are used for best-effort services, such as file downloads, email, and Internet browsing. 3GPP defines the series of standardized QCI listed in [8]-pag 52. The LTE standard provides drop mechanisms and priority management strategies in case of network congestion. To this purpose, each bearer is associated to an Allocation and Retention Priority (ARP). For example, ARP is used for the call admission control, to decide whether or not the requested bearer should be established in case of network congestion. However medical and healthcare services are often characterized by very particular and stringent requirements. In order to ensure the desired QoS,

specific mapping schemes have been proposed for e-health services on LTE [15]. On the other hand, a strict separation into multiple single flows may turn out to be inefficient, especially in case of contemporary transmission from multiple and heterogeneous information sources. For this reason, transmission solutions based on data flow aggregation at the application layer are also taken into account. The standard QoS architectures provide QoS support in per-flow basis, that is, all video packets receive equal QoS treatment from the MAC. This is insufficient for video streams that would benefit from per-packet QoS differentiation in order to ensure that the most important video packets (*e.g.*, SVC base layer) get transmitted also under limited transmission capacity.

### **1.1.3 Expected Innovations At The eNodeB (PHY Layer)**

The physical layer suffers from fading and other channel-induced impairments. Both Additive White Gaussian Noise (AWGN) and Rayleigh fading channels are considered. More specifically, thermal noise, channel fading, channel dispersion, path loss and shadowing are the main factors that can corrupt the transmitted signals. Various dispersive channel models used in the LTE standard are considered in the proposed investigations. In particular, I focus on the employment and design of various coding and modulation schemes for supporting reliable transmissions over various communication channels. In terms of coding, turbo codes and Low-density Parity-check (LDPC) codes are considered. In terms of modulation, classic M-ary Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM) schemes are taken into account when communicating over flat Rayleigh fading and AWGN channels. As multiple access techniques, LTE/LTE-A supports SC-FDMA schemes for uplink. Tab. 1.1 provides a summary of the use case.

UC1	Ambulance and Emergency Area
Key challenges	<ul style="list-style-type: none"> <li>• Provide dynamic adaptation of real-time, high-quality lossy data to the available network resources</li> <li>• Improve wireless network performance supporting medical data transmission</li> <li>• Support service differentiation for emergency services, provide QoS for emergency connections</li> <li>• Monitoring available networks and their available resources</li> </ul>
Main technical and functional requirements	<ul style="list-style-type: none"> <li>• Real-time transmission of high quality video data</li> <li>• Transmission of medical data</li> </ul>
Expected Innovations	<ul style="list-style-type: none"> <li>• Efficient SVC encoder/decoder</li> <li>• Dynamic adaptation module for high quality data transmission</li> <li>• Improved wireless radio transmission capabilities via new coding and transmission algorithms</li> <li>• QoS provisioning for emergency services/connections</li> </ul>

Performance evaluation criteria	<ul style="list-style-type: none"> <li>• QoS and QoE measurements over wireless networks</li> </ul>
---------------------------------	---

Table 1.1: UC1 summary.

## 1.2 Use Case 2: Emergency Area With Multiple Casualties

This use case is illustrated in Fig. 1.2 and considers emergency scenarios and accidents, where multiple patients are involved and where there is a need to transmit and exploit/fuse multiple data flows to/from the emergency area. The fusion operation can involve heterogeneous and static (or with very limited mobility) sources of information present in the emergency area: the various equipments can communicate wirelessly with the infrastructure (wireless infrastructure networks provided for example by LTE/LTE-A could fully or partially cover the emergency area). Extension of the infrastructure network can be realised by direct device communication and by some local/personal area (such as IEEE 802.15.4a, 802.11, LTE) or mesh (such as Wireless Distribution System (WDS), 802.11s) networks. The use case (UC-2) is characterized by a large number of video source: for example, video surveillance systems are usually deployed in large public places to remotely assure a constant area monitoring. Moreover, several wireless-connected cameras are thus covering the emergency areas (*i.e.*, random passengers are filming the emergency event or first-aiders are provided of fisheye panoramic cameras). The wireless cameras are equipped with some sensing technologies (GPS receivers, electronic gyroscope, compass, etc.), which allow them to automatically determine their position and orientation in space. Differently from UC-1, the ambulance is not on the move and act as *video source collector* gathering the generated video fluxes in the Emergency Area from medical devices and from the accessible ambient cameras. Thus, heterogeneous information (health conditions, videos, people/device positions, etc.) are collected, jointly exploited/fused to be delivered to and from the hospital: several advantages can be expected from their use. For example, during the rescue operations the person of the emergency medical services may require a higher visual quality of a particular area where an injured person is laying on the ground. Immediately the systems reacts to the operator request determining the area location and selecting the camera which best captures the desired area. Then, the corresponding video stream is sent with high priority to the hospital. As in UC-1, the prompt medical information delivery of a patient allows to speed up his transportation to the most appropriate hospital and to pre-arrange the emergency

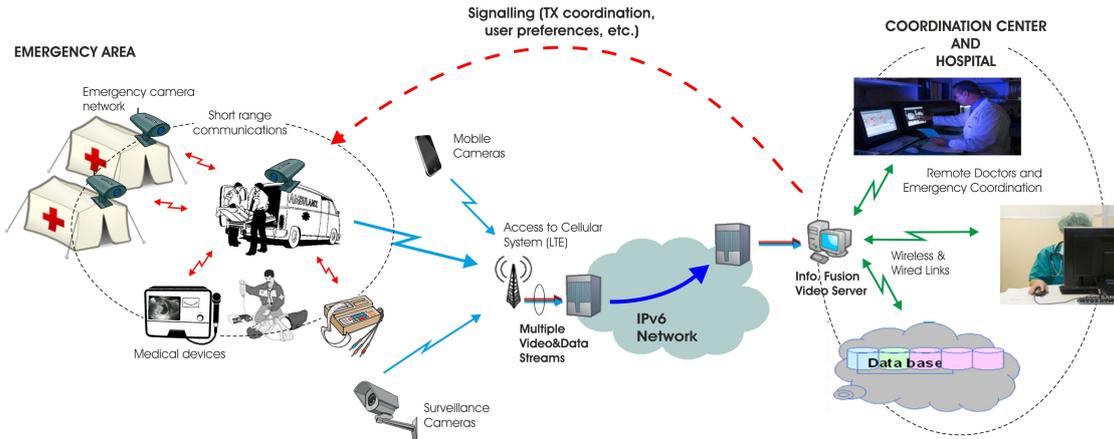


Figure 1.2: Emergency area with multiple casualties use case.

area at the hospital. Eventually, the person of the emergency medical services monitoring the situation can coordinate the first-aiders on the Emergency Area and/or rapidly understands that more medical staff is required and additional ambulances are sent to the emergency area. In this use case three different type of source are defined, in particular:

- ambient videos derived from the Emergency Area,
- ambient videos acquired on the ambulance,
- medical video and data information.

It is clearly noticeable that the different source of information present particular characteristics and they may cover different importance for a medical point of view. Ambient videos from Emergency Area (EA) (also called outdoor videos in the following) are fundamental for the remote operators to better understanding the emergency situation; however, videos and medical information from the ambulance present higher importance respect to the outdoor one since they are related to a specific patient condition. Differently from UC-1, I add up at the SVC video adaptation the source prioritization based on the video importance. The presence of multiple sources of information (*e.g.*, medical devices, cameras, and smartphones) requires some coordination in the communications to/from the hospital, especially in case of limited availability of transmission resources. To this purpose, several criteria can be adopted to determine transmission policy and priority, such as perceived video quality, critical events related to the state of health of patients, position of injured people in the emergency area, monitoring capabilities of devices, wireless resources availability, etc. [7]. Due to these reasons, a

Coordination Center, which can be an external center or a dedicated room inside the hospital, is also included.

### **1.2.1 Expected Innovation At The Coordination Center (Application Layer)**

The Coordination Center (CC) is developed with the aim to manage the multiple sources of information destined to end-users located in the hospital; as it is shown in Fig. 1.2, it acts as the main interface between the hospital and the outer world. In this regard, different functionalities will be implemented in this module. The CC is responsible to coordinate all the received data and, taking into account the requests of the doctors, can prioritize different fluxes of data or ask to the senders in the emergency area to adapt their transmissions, the compression of the sent data and the kind of data transmitted. Then, the received information are forwarded, based on specific selection criteria, to the destination user. Outdoor videos can be generated by a redundant number of cameras, considering large monitored areas, spatially placed in various positions. The large number of involved devices may guarantee a global monitoring of the area of interest. However, before each transmission windows it is necessary to sort the outdoor ambient cameras in order to select only a specific camera sub-set. This necessity comes out for the sake of optimizing the usage of distributed hardware resources and the usage of the radio resources. Moreover, the camera sub-set selection has to be performed taking into consideration specific quality requirements to satisfy the QoE of the final users. Due to these reason, several camera ranking techniques will be implemented within the CC.

### **1.2.2 Expected Innovation At The eNodeB (DL And MAC Layers)**

Expected Innovation At The eNodeB (DL And MAC Layers)

Within the DL framework a specific functionality at the LTE/LTE-A eNodeB is included, in charge of managing multiuser scheduling for uplink, allocating the most appropriate radio resources (denoted as Physical Resource Blocks (PRBs)) capable to guarantee the required QoS and selecting the physical transmission schemes (modulation and coding). The radio channel capacity or transmission rate, may vary in time from PRB to PRB and from user to user. A packet-scheduling mechanism must then be used at the eNodeB to establish the frame-to-frame resource allocations to ensure equitable treatment of all data users, according to individual traffic type/QoS specifications and channel transmission conditions. The purpose of scheduling is two-fold: to provide the appropriate QoS measures such

as maximum delay, minimum rate, maximum packet loss probability, and other QoS performance guarantees to individual users, where the propagation conditions allow this, and to ensure full and efficient use of the resources, most commonly link bandwidth or capacity. Scheduling or resource allocation algorithms are often designed to achieve some desired performance objectives, through the maximization of suitably defined utility functions. These objectives could include the attainment of a desired QoS for each data traffic type. They would also include the attainment of maximum possible overall throughput on a given radio link, as stated above. This requires a cross-layer framework, which makes some input information available to the scheduler: channel state information or channel quality indicators from the physical layer and source priority and ranking information, target source rate, target link quality from application and upper layers. For more advanced cross-layer functionalities, such as the joint source adaptation and radio resource allocation which may be considered to optimize video delivery over the network, the resource allocation unit at the air interface can send information, like available rate and delays for each service queue, to upper layers.

### **1.2.3 Expected Innovation In The Ambulance (Application Layer)**

The processing unit on-board the ambulance is responsible to collect and aggregate the multimedia streams from all the cameras available in the emergency area and to transmit them through the LTE link, toward the CC/hospital. To perform its processing, the ambulance exploits the ranking information received from the CC and the achievable throughput across the LTE link. The target video coding rates of the selected sources is jointly adapted to guarantee the optimal level of quality (in terms of SSIM) compatible with the available radio resources and communicated to the application controller of each encoder. Adaptation can also be decided directly by the senders according to the cross layer information received, while not badly affecting the overall network efficiency. Different devices can have more or less intelligence and adaptation capabilities. In this use case, the ambulance, which relays both context and medical data, can implement more advanced adaptation techniques with respect to the smartphone or the ECG device, such as, for example, SVC encoding of data received by different sources. This also leads to an improved utilization of the network resources, as shared possibly by other critical applications. Tab. 1.2 provides a summary of the use case.

<p><b>Use case 2</b></p> <p>Key challenges</p>	<p><b>Emergency area with multiple casualties</b></p> <ul style="list-style-type: none"> <li>• Real-time management and source adaptation for multi-view transmission</li> <li>• Content-aware transmission of heterogeneous information (video, medical data, etc.) over LTE radio links</li> <li>• Automatic and/or user-driven camera selection based on heterogeneous context-information (position, resolution, etc.)</li> </ul>
<p>Main technical and functional requirements</p>	<ul style="list-style-type: none"> <li>• Medium bit-rate per camera</li> <li>• Cellular and broadband access technologies (such as LTE) to enable communications between the EA and the hospital</li> <li>• Self/mutual localization capabilities of deployed cameras and devices in the emergency area</li> <li>• Automatic switching of video sources and stream adaptation according to the available transmission resources, user preferences and video quality criteria</li> <li>• Multi-view video of the patient</li> </ul>

Main technical and functional requirements	<ul style="list-style-type: none"> <li>● Multi-source information exploitation/fusion and coordinated transmission based on critical events and/or remote user preferences</li> </ul>
Expected Innovations	<ul style="list-style-type: none"> <li>● Multi-view exploitation/fusion techniques and source adaptation</li> <li>● User or event-driven adaptation and coordinated transmission</li> <li>● Scheduling and radio resource allocation for video transmission in LTE uplink</li> <li>● Cross-layer optimization techniques for efficient video and medical data transmission with adequate QoE, according to available resources and network capability</li> <li>● Novel cross-layer scheduler in the network, with the aim to support a proportional model for QoS independently from a specifically deployed wireless technology and actual radio channel status.</li> </ul>

Performance evaluation criteria	<ul style="list-style-type: none"> <li>• Video quality metrics (PSNR, SSIM, PMOS)</li> <li>• Achievable throughput and delay</li> <li>• Capacity of the source adaptation and multi-view management algorithm to react to critical events and sudden context changes</li> </ul>
---------------------------------	---

Table 1.2: Use case 2 summary.



# Chapter 2

## Camera Selection Techniques For Large Camera Networks

### 2.1 Introduction And State-of-the-art

In recent years the study of Visual Sensor Networks (VSNs) has become a very attractive field for the scientific research. VSNs are characterized by huge capabilities and they support a great number of innovative applications. In particular the VSNs are employed for remote monitoring applications where direct visual information are essential for controlling and managing particular events [16, 17]. VSNs are mainly formed by visual sensor nodes more commonly called video camera nodes. Each node integrates the image sensor and some embedded processing units. In such applications where cameras have fixed and easy-to-reach positions they can be connected via cable; in other cases they are equipped with a wireless radio interface [9]. VSNs are a useful tool in a number of social, research, educational applications, including:

- *surveillance*: protection of large area (airports, subways, stadiums, etc.) where detection and tracking mechanisms are required to localize possible intruders. This is one of the primary applications of VSNs, with a great number of cameras disseminated around the area of interest [18].
- *environmental monitoring*: video-cameras are deployed to monitor inaccessible or wild areas with the aim to acquire environmental images of the scene. Typical examples of such applications are: disaster sites management, traffic control in highways and natural environments monitoring such as forest and deserts.
- *telepresence and remote visiting systems*: the usage of multiple cameras permits a remote user to visit public buildings such as museums and art galleries

and to participate to video conferences. Sometimes the underlying camera networks are developed to provide a user with visual information with arbitrary viewpoint. In this way the user can explore the scenario of interest "moving" inside of it and changing the angle of visualization.

- *teleconsultation and distance healthcare*: in this case the VSNs constitute a useful instrument to aid remote medical assistance and to support clinical therapy from a distance. This is especially important for people with limited access to the healthcare system: people with disabilities, individuals living in remote rural areas or emergency scenarios or involved in accidents, with multiple injured people.

### 2.1.1 Video Transmissions In Multi-Camera Systems

One of the most critical issues concerning the VSNs consists in the large amount of data that each camera collects from the monitored environment. Differently from typical sensors (i.e. temperature or pressure sensors that provide measurements of unidimensional physical data), each image sensor collects a bi-dimensional set of data representing a captured image. The nature of the collected data entails complex tasks of information analysis and processing, as well as huge amount of data to be transmitted through the network. Moreover the majority of applications require real-time transmissions and this means strict requirements concerning the maximum allowable delay. For example the problem of the channel bandwidth optimization in the context of VSNs has been addressed in [19]. In that work a minimum bandwidth, for each selected camera, was guaranteed. Furthermore the remaining available resources were dynamically allocated considering the camera selection policy and the video contents.

### 2.1.2 Embedded Processing

Based on the previous considerations, the optimal trade-off between the information accuracy and the network efficiency is to be pursued. To this purpose the embedded camera processing assumes an important role. The objective of the local processing consists in reducing the data flows to transmit or in providing some information about the context in order to assist the selection of proper acquisition and transmission policies. The local processing can be performed with different levels of intelligence, depending on the context on which the camera system is deployed. The most common processing operations include: background subtraction for motion/object detection [18],[20], edge detection and more complex algorithms such as feature extraction, object classification and scene reasoning [21, 22].

### 2.1.3 Sensor Collaboration

In some specific applications, such as detection and tracking of targets, VSNs need to cover wide areas and require highly detailed information. For this reason, the camera network manages a great number of information flows derived from a redundant number of interconnected nodes. The overall quality information can be increased by investigating the correlations between the individual camera data flows. A typical example is when two or more cameras have partially or completely overlapped Field of Views (FoVs). In this case the collaboration between sensors is defined as *spatial-based*. Many advantages can be obtained by spatial collaboration: for instance, through cross-validation, the reliability of the detection process is incremented. Furthermore some triangulation techniques can be applied permitting an accurate description of the target object (in terms of position and velocity) [23]. Even if two cameras are oriented in different directions, they can still cooperate. For example, in [24] a distributed scheme for a target tracking system has been studied. Given the position and the orientation of the nodes, a camera subset is dynamically selected for the object detection, while the remaining neighbouring cameras are used to extract and share the target features, which are correlated across the nodes.

### 2.1.4 Sensor Management And Camera Sselection

In the context of large camera networks, composed by a redundant number of camera nodes, the camera selection and management play a fundamental role. Sensor management defines the policies for scheduling the camera nodes activity. In each moment a subset of cameras has to be selected in order to perform continuous monitoring and collect information with the required quality. The purpose is to provide visual information able to satisfy the specified application requirements. Several cost metrics for the camera selection are investigated in the literature. In particular, the camera selection is of primary importance if the network is composed by self-supplied wireless camera nodes. In this case each camera is managed with the aim of defining the proper task and the activity period. The selected subset of involved cameras can be changed over time in order to balance the camera energy consumption. In [25] the authors propose a method to manage the power consumption of camera nodes. In the proposed solution each node is awake for a certain period of time and the camera node enters a low-power configuration based on the time-out status of its neighbouring cameras. Again, geometrical considerations about the placement and the orientation of the nodes are taken into account to defined further camera selection cost metrics. Each camera node is characterized by a directional sensing model: video cameras capture objects/scene images placed in different spatial positions. The gathered information are depen-

dent on the direction on which the camera is oriented and the 3D viewing volume defined by the camera FoV. Several works in the literature analyze these aspects. In [26] a camera selection algorithm is provided with the aim to reconstruct a view from a user-specified view point. In particular two cost metrics are defined: the coverage-aware solution considers the remaining energy of the nodes and the overall coverage of the indoor space, while the quality-aware favours the selection of cameras with the most similar angle of view with respect to the user viewpoint. The video quality information also depends on the placement of the cameras. The optimal placement of the camera nodes in use is also addressed in recent works. An example is reported in [20], in which a good camera-network deployment is proposed mainly based on the target position. In its turn, [27] provides a model to manage a set of cameras that is basically focused on the occupancy of the monitored area. Only a limited number of cameras are involved and they send a limited amount of data in accordance to the energy-saving requirements and the bandwidth constraints. the final goal consists of minimizing the area potentially occupied by the moving objects and finding the number of cameras necessary to provide a sufficient visual hull area for each object.

### **2.1.5 A Preliminary Camera Selection Technique For Healthcare And Safety Applications In Emergency Areas**

In the context of healthcare and safety, multiple camera systems can support a great number of novel video-based services. In particular, some challenging applications that can be performed thanks to VSNs include real-time consulting and diagnosis from a remote location, transmission of multimedia contents from emergency scenarios and remote surgical assistance. The potential benefits of using camera network systems are manifold and include:

1. the possibility to visit more patients in a limited amount of time,
2. the possibility to have an overall control of the involved injured people and to coordinate the work of the first-aid personnel in emergency areas,
3. the opportunity to select the most suitable viewpoint in tele-diagnosis and remote medical assistance.

Moreover a wireless camera network, developed for this kind of scenarios, has to satisfy most of the requirements and constraints mentioned in the previous paragraph. For example a reasonable placement of the nodes has to be guaranteed in order to monitor the whole area of interest. Furthermore a redundant deployment of the nodes assures the advantages of spatial collaboration and provides useful information even if some cameras are not available or occluded. In addition, the

camera selection policy has to be designed with the aim to collect information with a high level of accuracy. Only few works in this research field specifically address the image/video quality perceived by the final user as a main criterion for the camera selection problem. On the contrary, in medical applications the QoE for the remote user is of crucial importance. For example, in [28] the authors choose the camera subset that best provides the quality of the reconstructed camera view. The basic idea proposed in [28] is to divide the 3D viewing volume (frustum) of each camera in smaller unit volume (sub-frustum). Then the selection criterion is based on the fact that the farther an object is from the centre of a sub-frustum the lower will be the quality of its representation in the captured images. Thus, based on the distance of each point from the sub-frustum centre, they select the camera that can display this point with the minimum visualization error. In [28] the concept of quality is strictly correlated to the position of the object of interest.

## 2.2 Camera Ranking Techniques

In this section, two novel content and context-aware camera ranking algorithms are proposed. Differently from the works present in literature and for the best of our knowledge, I approach the problem of the camera ranking with the aim to maximize the QoE by the final user. The proposed ranking strategies are designed taking into account some intrinsic parameters, the resolution and the frame-rate which are characterizing factors for each camera. Furthermore, the camera selections are performed taking into account camera positions and the related distance from a given monitored object. Two distinct algorithms are developed in order to be adaptively applied to the context in which the camera network is deployed. The first algorithm is defined as Camera Ranking for Static Scenes (CRSS), it requires a lower computational burden and is oriented for static or low motion scenes [1]. The second algorithm considers also the quantity of motion that is captured by each camera. Due to this reason, the proposed scheme will be denoted as Camera Ranking for Dynamic Scenes (CRDS) in the following [2]. In Tab. 2.1 the ranking criteria taken into account by the two proposed algorithms are provided. To evaluate the effectiveness of CRSS and CRDS, different sets of simulations are performed using cameras that are characterized by different spatial and temporal resolution. At the end of each simulation, the proposed algorithms will be validated through objective metrics such as PSNR and SSIM. Finally, the proposed algorithms will be further validated through subjective tests, such as PMOS. The remainder of this chapter is organized as follows: in Sec. 2.2.1 a realistic camera ranking scenario is presented in which the proposed solutions are applied. Then, the proposed camera ranking techniques are presented in detail in Sec. 2.2.2 and in Sec. 2.2.3 respectively. In Sec. 2.2.4 preliminary tests over the CRSS for emergency

	Parameters used in the algorithms	CRSS	CRDS
Camera Info	FoV	x	x
	Resolution	x	x
	Frame-rate		x
Position Info	Camera Position	x	x
	Position of the Scene of Interest	x	x
	Orientation	x	x
Motion Info	Object Velocity in the Scene		x

Table 2.1: The ranking criteria taken into account by the two proposed algorithms.

areas and medical contexts are presented, while in Sec. 2.2.5, technical details are provided about the procedures I apply to extract specific camera parameters (*i.e.*, the intrinsic matrix) and the captured motion quantity by each camera. Finally, numerical results are presented regarding the camera ranking algorithm comparison, the validation through the objective metrics and the validation thanks to some subjective test sessions that will give us the opportunity to verify the algorithm capability in satisfying the QoE.

### 2.2.1 Scenario Description

Let us consider the camera network scenario in Fig. 2.1, where the target (patient at the center of the figure) has to be monitored through the usage of a multi-camera system. Different wireless cameras with monitoring functionalities are placed in different positions. Each camera covers a particular portion of the considered area depending on its orientation and FoV. In Fig. 2.1,  $\theta$  corresponds to FoV/2 and the optical center of Camera A is highlighted. The knowledge of the point of interest allows a preliminary camera selection discarding all the cameras not capable to provide useful information for the final user. For example, camera B will be not taken into consideration in the final ranking. This scenario is included in the use case presented in Sec. 1.2. Some assumptions could be made to facilitate the description of the camera network systems. First, a pinhole model, which is the most widely used for mapping a 3D scene into a 2D image, has been adopted to describe the camera nodes [29]. For a more realistic model, the pinhole is enriched with the camera FoV, which delimits the portion of the 3D monitored area. In Fig. 2.2-A a 3D pinhole camera model has been depicted, with different vertical and horizontal FoVs. In the figure, the camera optical center is  $\mathbf{c}$ , the point to be monitored is  $\mathbf{p}$ , and  $\mathbf{u}$  determines the camera orientation. Based on the camera pose and the point of interest, I delimit the Area of Interest (AoI) as illustrated in Fig. 2.2. The AoI is the 2D portion of the real-world in which the scene of interest is located. Considering the linear dependence between pixels and visualized area (in

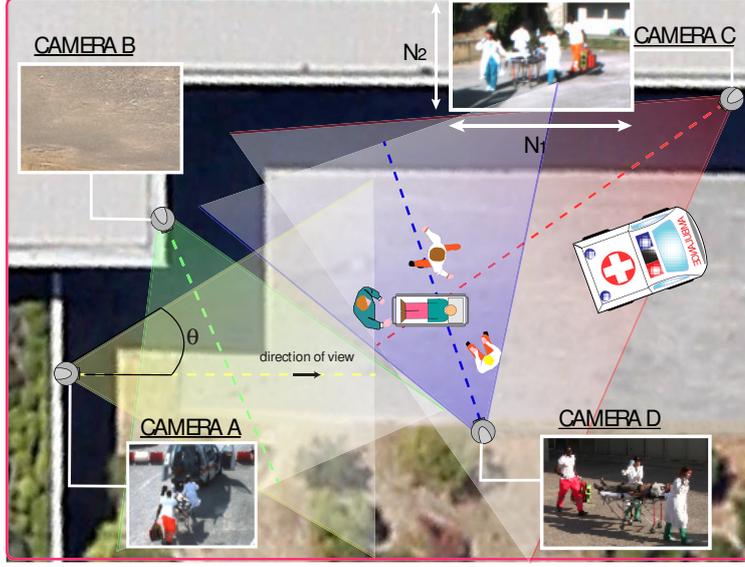


Figure 2.1: Example of VSN.

square meters), the proposed 3D camera model is decomposed in two 2D models where the number of linear pixels per meter is computed considering each FoV singularly. In addition, assuming the camera model has no rotations with respect to the global system reference, a geometric transformation is applied to translate the global system reference into the camera system reference. Thus, in Fig. 2.2-B,  $\mathbf{p}' = \mathbf{p} - \mathbf{c}$  denotes the target position;  $\mathbf{u}' = \mathbf{u} - \mathbf{c}$  describes the camera orientation;  $d$  is the distance between the optical center and the monitored target; finally,  $N_1$  is the number of linear pixels along the first dimension of the image plane. The angle between the orientation vector  $\mathbf{u}'$  and the target position  $\mathbf{p}'$  is

$$\alpha = \arccos \left( \frac{\mathbf{p}' \cdot \mathbf{u}'}{|\mathbf{p}'| \cdot |\mathbf{u}'|} \right). \quad (2.1)$$

The projection of the distance  $d = \|\mathbf{p}'\|$  on the optical axis is calculated as

$$d' = d \cos \alpha [m] \quad (2.2)$$

and the length visualized along the considered image plane is

$$M = 2 d' \tan \theta. \quad (2.3)$$

### 2.2.2 Camera Ranking for Static Scenes (CRSS)

The proposed ranking algorithm was based on geometrical concepts and takes into account some camera intrinsic parameters and the camera resolution. The main

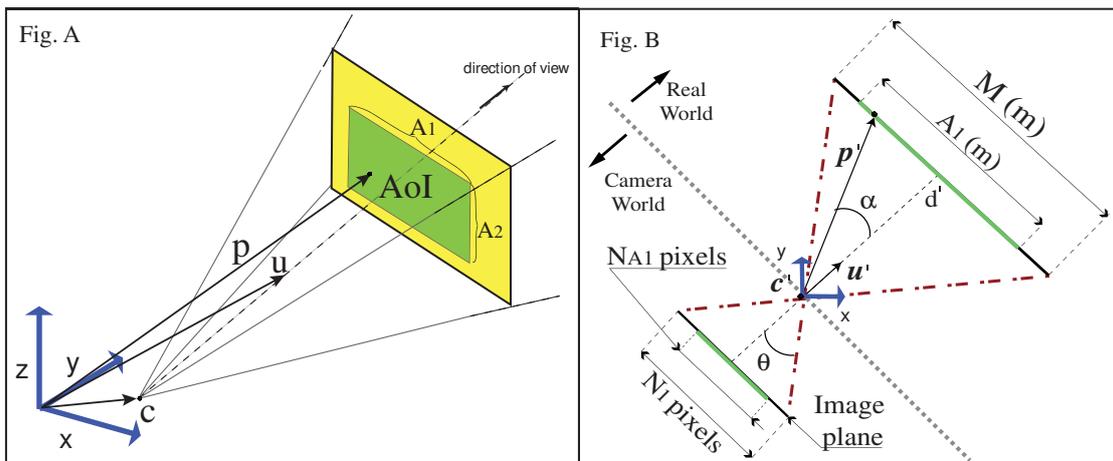


Figure 2.2: The pinhole camera model: A) 3D representation; B) 2D representation.

purpose of the algorithm is to find the camera sub-set able to provide the best image quality in low-motion scenes. The considered camera parameters in the ranking process are:

- the pose (position and orientation) of each camera,
- the resolution of each camera.

Then, the principle behind the CRSS is focused on the capability to combine the size of the object and the pixel density on the image plane. The object size projected on the image plane depends on the distance of the object to the camera: the closer are the objects, the bigger are their projections on the image plane. Moreover, the pixel density depends on the resolution of the single camera: the higher is the resolution, the more accurate is the resulting image. The ranking that is presented in [1] consists in calculating the amount of pixels required to represent a single squared meter surface at the distance which the target is located. To this purpose, the number of linear pixel per meter is obtained in the following way:

$$q_{lin} = \frac{N_1}{M} [ppm]. \quad (2.4)$$

Clearly, assuming a squared aspect ratio for the pixels, the number of pixels per unit area at the considered distance is  $q = q_{lin}^2$ . The greater is the  $q$ -value of a camera, the higher will be the final position in the ranking. Additionally, the number of pixels  $N_A$  corresponding to the projection of the AoI on the image plane is calculated. Given the AoI, the resulting resolution is obtained as follows:

$$N_{AoI} = N_{A1} \cdot N_{A2} = q \cdot AoI = q(A_1 \cdot A_2). \quad (2.5)$$

### 2.2.3 Camera Ranking for Dynamic Scenes (CRDS)

In this section a novel ranking strategy, called Camera Ranking for Dynamic Scenes (CRDS), is proposed. Conceptually, the CRDS takes into account the spatial features considered in CRSS, but includes in the ranking criteria also the acquisition frame rate and the average motion feature velocity. In order to study the impact of frame rate and resolution on the final perceived video quality, it is assumed without loss of generality that each video sequence can be modelled as a three-dimensional continuous signal. The video signal is acquired by the cameras and then is discretized taking into account the frame rate and the spatial resolution as sampling frequencies. I start by defining a model for the intrinsic video power spectrum of the filmed scene. Then, the camera quality for the specific scene is estimated as the amount of intrinsic video power that the camera can acquire. Given the sampling frequencies, this acquired power depends on the signal bandwidth in the spatial and temporal domains: the larger is the sampling frequency, the greater is the collected power by the signal. The spatial sampling rate depends on the geometry of the scenario and on the scaled resolution, based on the CRSS principle. The role of the temporal sampling rate depends on the scaled video motion. In this regard, we used the analysis leading to the concept of spatio-temporal power video spectrum, as proposed in [30].

#### Natural Time-Varying Images Power Spectrum Definition

In [30], the authors started by analyzing a thousands of video segments, in order to identify common regularities in natural scenes. A video segment is a 3-dimensional hypercube of size  $L \times L \times T$ , where the first two dimensions refer to the spatial coordinates of a point in a frame, while the third dimension identifies the time instant within the sequence. Indicating with  $s(\mathbf{x}, t)$  the windowed light intensity in point  $\mathbf{x} = (x_1, x_2)$  at time  $t$ , the correlation between two points separated by the spatio-temporal distance  $(\boldsymbol{\xi}, \tau)$  can be expressed as [30]:

$$r(\boldsymbol{\xi}, \tau) = \frac{1}{L^2 T} \int_0^L \int_0^L \int_0^T s(\mathbf{x} + \boldsymbol{\xi}, t + \tau) \cdot s(\mathbf{x}, t) dx_1 dx_2 dt \quad (2.6)$$

where the spatial displacement is  $\boldsymbol{\xi} = (\xi_1, \xi_2)$ . The power spectrum of the considered spatio-temporal segment is obtained through the Fourier transform of  $r(\boldsymbol{\xi}, \tau)$  [30]:

$$R(\mathbf{f}, w) = \int_{-L}^L \int_{-L}^L \int_{-T}^T r(\boldsymbol{\xi}, \tau) \cdot e^{j2\pi(\mathbf{f} \cdot \boldsymbol{\xi} + w\tau)} d\xi_1 d\xi_2 d\tau \quad (2.7)$$

where  $\mathbf{f} = (f_1, f_2)$  are the spatial frequencies and  $w$  the temporal frequency, respectively. Assuming that the objects in the scene are placed at a distance in

the range  $[d_1, d_2]$  from the camera and that their static spectrum is rotationally symmetric, the following expression is derived for the average power spectrum [30]:

$$G(f, w) = \frac{K}{f^{m+1}} \int_{d_1}^{d_2} P\left(\frac{w}{f}z\right) z dz \quad (2.8)$$

where  $f = \|\mathbf{f}\|$ ,  $P(\cdot)$  is the velocity distribution of the objects in the scene along a certain direction<sup>1</sup>, and  $K$  and  $m$  are two parameters whose values are estimated numerically. Note that in (2.8) the power spectrum depends on the velocity distribution, which in many natural scenes can be approximated by simple power-law distribution [30]. Finally, the power spectrum of natural varying images is modeled as [30]:

$$G(f, w) = \frac{K\bar{v}}{2f^{m-1}w^2} \left[ \frac{n-2}{(x+1)^{n-1}} - \frac{n-1}{(x+1)^{n-2}} \right] \frac{wd_2}{fv_0} \frac{wd_1}{fv_0} \quad (2.9)$$

where  $\bar{v}$  is the average object velocity, and  $v_0$ ,  $n$  are constant values. This analytical model was numerically validated in [30] by observing that the measured power spectrum of the considered video segments, once scaled by the factor  $f^{m+1}$  and expressed as a function of  $f/w$ , shows a behavior in total accordance to (2.9). Hence, the total collected power can be calculated as:

$$P = \int_0^{w_{ul}} \int_0^{f_{ul}} G(f, w) df dw. \quad (2.10)$$

For each camera,  $f_{ul}$  and  $w_{ul}$  depend on the spatial and temporal bandwidth for the considered video signal. According to the Shannon sampling theorem, the camera frame rate  $F_r$  acts as a temporal sampling factor transforming the continuous signal in a discretized version. Thus, the temporal bandwidth is  $w_{ul} = F_r/2$ . To determine the spatial frequency, we should take into account (see Fig. 2.2-B) the length visualized along the considered image plane  $M$ , given by (2.3), the distance between the camera optical center and the filmed scene  $d'$ , and the number of pixels forming the AoI on the image plane,  $N_{A1}$ . The spatial sampling frequency in cycles per degree is therefore

$$S_f = \arctan\left(\frac{M}{d' \cdot N_{A1}}\right)^{-1} \text{ [cycles/degree]} \quad (2.11)$$

To determine the spatial frequency in (2.11) I take into account: the length in meters visualized along the considered image plane  $M$  (2.3), the distance between the camera optical center and the filmed scene  $d'$  (see Fig. 2.2-B) and  $N_{A1}$  that is

---

<sup>1</sup>Note that in the model of [30] also the velocity distribution of the objects in the scene is assumed as rotationally invariant.

defined in (2.5) and corresponds to the number of pixels forming the AoI on the image plane. In (2.11),  $b = M/N_{A1}$  defines the length in meters described by each singular linear pixel while  $b/d'$  provides the ratio between meter per linear pixels and distance between camera and AoI. Moreover, (2.11) expresses the spatial frequency in cycles per degree. The principle behind the formulation in (2.11) is based on the fact that the spatial frequency jointly depends on the distance from the AoI and the spatial resolution  $N_{A1}$  (or  $N_{A2}$ ). The best spatial frequency is guaranteed when the AoI fullfill the whole image plane: more precisely, when the camera is placed at the distance such that  $N_1 = N_{A1}$  (Fig. 2.2-B) is satisfied. In this ideal case the scene of interest is sampled taking into account the whole spatial resolution of the considered camera. Increasing the distance the projection of the AoI on the image plane gets smaller dimension: respect to the ideal case, the spatial resolution  $N_{A1}$  is lower; it follows that the spatial sampling is coarser given a resulting lower spatial frequency. For the sake of clarity, cases where AoI exceeds the image plane are not considered. where  $b = M/N_{A1}$  is the length projected on one linear pixel. Once the spatial and temporal bandwidth have been calculated for each camera, the CRDS uses (2.10) for ranking: the higher is the power, the higher will be the camera classification in the ranking.

## 2.2.4 Preliminary Tests Over The CRSS For Emergency Areas And Medical Contexts

In this test session, the CRSS is applied to some video flows that are acquired from a real camera network. The objective is to validate and confirm the theoretical efficiency of the proposed algorithm. To this aim, the camera network is initially designed. An area to be monitored is defined and a set of cameras is placed in different positions and with different orientation [10]. The camera placement has been designed in order to cover the whole area and, in some parts, two or more cameras partially monitor the same sub-area in a redundant way. Furthermore, the camera network is assigned to monitor an object on the move (in the following it is also defined as the target to be monitored) within the area of interest. In each moment, different camera subsets are capable to monitor the target depending on its position. In Tab. 2.2 all the cameras are listed (in the C,D,F and G cases the device embedded cameras) for the testing phase. It is noticeable that cameras have different FoVs (h stands for horizontal and v for vertical) and resolution. The FoVs in conjunction with the position and the orientation permit to define the portion of the 3D area that the different cameras can monitor: only if the target lying inside the delimited 3D area can be visualized on the specular image of a camera (Fig. 2.3-(a)).

The developed algorithm is applied to the described scenario in order to:

	Cameras	FoV <sub>h</sub> (rad)	FoV <sub>v</sub> (rad)	Resolution
A	Canon IXUS 950IS	0.969	0.7520	640x480
B	Canon IXUS 105	1.1362	0.8930	640x480
C	Acer Aspire 5738Z	0.9339	0.5527	640x360
D	iPad 2	0.7657	0.4455	1280x720
E	Canon Legria HRF106	0.6641	0.3832	1920x1080
F	Motorola RAZR XT910	0.8727	0.5130	1920x1080
G	Nokia E71	1.0015	0.7790	320x240

Table 2.2: Set of cameras used during the test phase. In evidence the measured FoVs and the resolution.

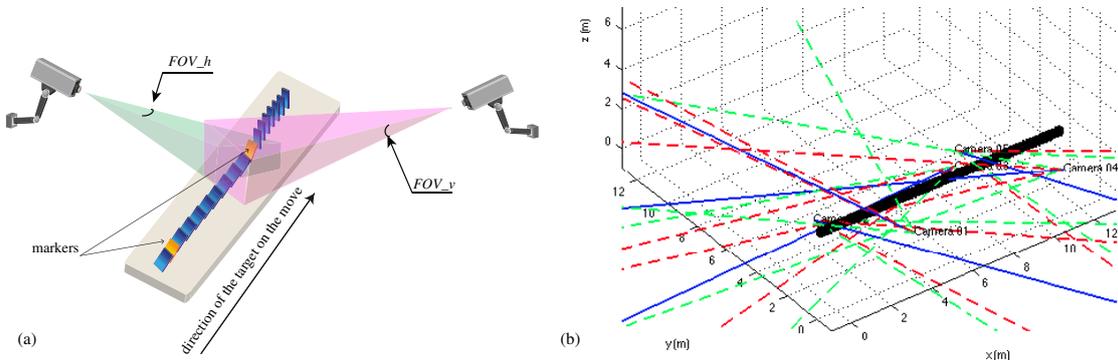


Figure 2.3: (a) Camera network composed by two nodes. (b) 3D representation of the camera network designed for the test session.

- detect which cameras are contemporary recording the target,
- perform a ranking on the detected camera subset,
- select the camera that provide the best quality considering position, orientation and resolution as ranking parameters.

After the camera deployment, the presence of an object on the move is simulated within the monitored area. It is remarked here that this is a preliminary test phase. The main goal is to apply our camera ranking strategy on real videos and to evaluate the algorithm efficiency. According to this reason, an easy-to-reproduce indoor scene is chosen. In particular, the camera network capture the video sequence of a domino fall down scene: the target on the move is represented by the fall front of the domino tiles. It is evident from Fig. 2.3-(a) and Fig. 2.4 that the target followed a rectilinear trajectory. It is also remarked that further tests are currently ongoing based on the camera selection algorithms in simulated

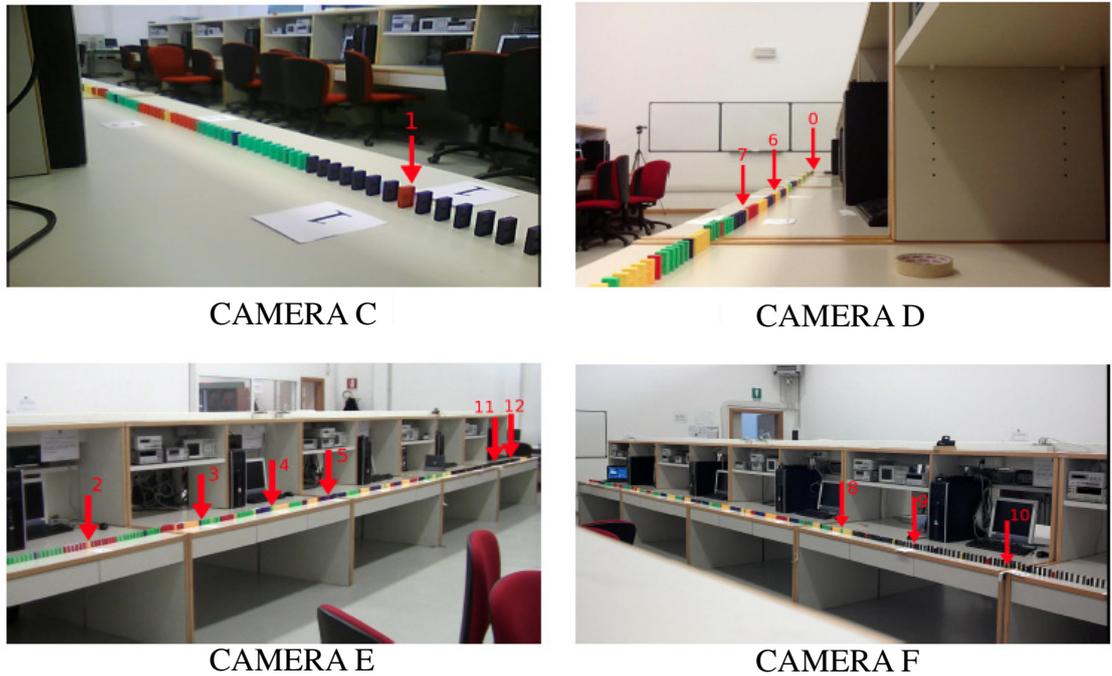
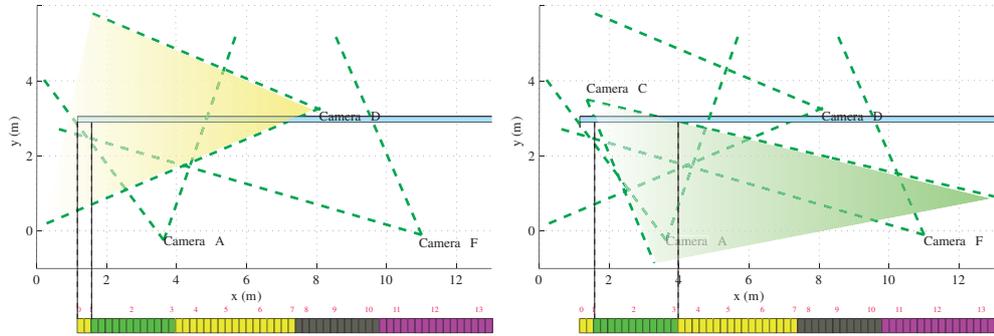


Figure 2.4: Frames acquired by the selected cameras. In evidence the differentiation of the target position on which the camera selection is computed.

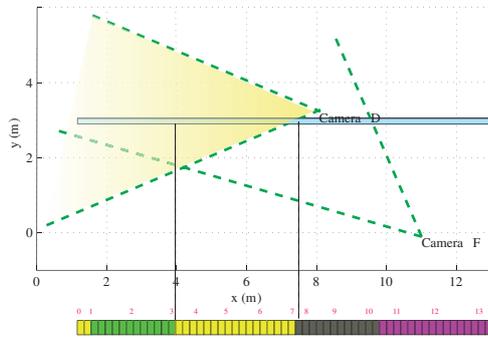
emergency contexts. Some target positions are selected (they are recognizable by the markers in Fig. 2.3-(a) and Fig. 2.4) which constitute some reference points on which the proposed camera selection technique is applied. For each of the different selected position, the camera ranking algorithm is computed in order to find the camera with the best visualized quality. In Fig. 2.3-(b) the 3D representation of the camera network and the monitored area (in this particular example, the dominoes line) have been illustrated. For each node, the vertical FoV (red dotted line), the horizontal FoV (green dotted line) and camera orientation (blue continuous line) are depicted, while FoVs amplitude and resolution are listed in Tab. 2.2. For the sake of clarity, through an example, which camera is selected for each different target position is showed. In Fig. 2.5 the 2D representation of the camera network is presented. In the mentioned figure thirteen different target positions are enumerated and only the horizontal FoV is depicted for the cameras. The first part of the domino line is covered by the camera subset A and D (Fig. 2.5-a). The camera selected, by the ranking algorithm, is D despite the fact that is the farthest from the target. In fact, the camera A is placed at a minor distance from the target, but the higher resolution of D entails a better quality. In the second part (Fig. 2.5-b), the competition concerns cameras A,C,D and F. The selected

camera is C. The node C is the nearest to target and can optimize the visualized quality despite the lower resolution. In its turn, the central part (Fig. 2.5-c) is captured by camera D given the proximity to the target and the high resolution in comparison to the other cameras. Again, the sub-area determined by markers 8,9,10 (Fig. 2.5-d ) is monitored by the camera F. In this case, the mentioned area is commonly covered by F and, in part, by G. However, only F is selected: the highest resolution of F overcomes the major proximity of G to the target. To conclude, the camera selection recognizes the camera G as the only one capable to capture the two final positions of the target (Fig. 2.5-e) and it is selected despite the lowest resolution. In addition, in Fig. 2.6 the acquired video frames by the selected cameras is proposed. In particular, each couple of sub-figures is linked to the initial and the final target position captured by the different cameras. It is noticeable that the final target position of a camera corresponds to the initial position of the next selected one. It proves that the camera ranking strategy detects which cameras are capable to monitor a specific area. Moreover, the camera candidate is dynamically selected on varying the target position. In this way, the object of interest is constantly monitored by the camera that provides the best QoE in terms of visualized quality.

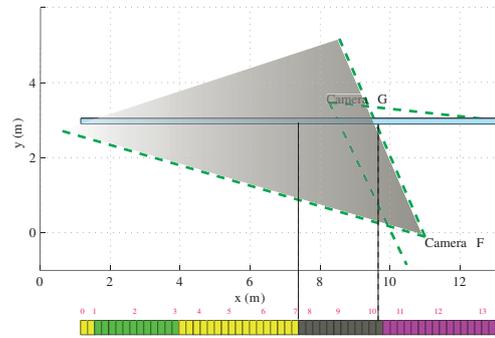


(a) CAMERA D

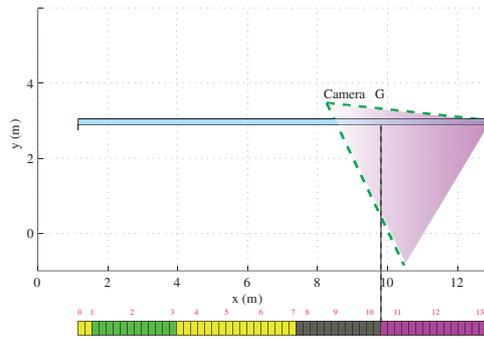
(b) CAMERA C



(c) CAMERA D



(d) CAMERA F

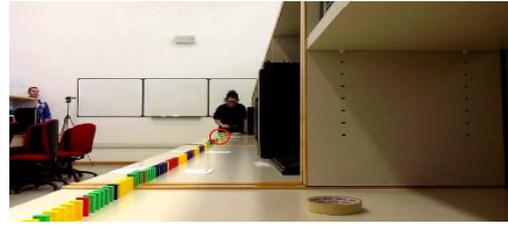


(e) CAMERA G

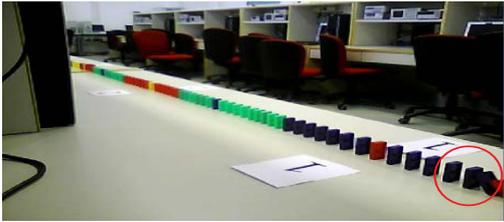
Figure 2.5: The camera selection result for the different target positions.



(a) CAMERA D, Target Position: 0



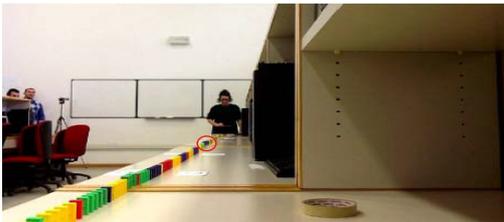
(b) CAMERA D, Target Position: 1.



(c) CAMERA C, Target Position: 1.



(d) CAMERA C, Target Position: 3.



(e) CAMERA D, Target Position: 3.



(f) CAMERA D, Target Position: 7.



(g) CAMERA F, Target Position: 7.



(h) CAMERA F, Target Position: 10.



(i) CAMERA G, Target Position: 10.



(j) CAMERA G, Target Position: 13.

Figure 2.6: Video frames acquired by the selected cameras based on the target position. For each camera, the initial and the final captured target position is highlighted.

Camera ( $C_i$ )	Image dim.	AoI dim.	Frame rate ( $F_r$ )
1	720×576	320×320	25
2	720×576	320×320	12.5
3	720×576	320×320	6.25
4	360×288	160×160	25
5	360×288	160×160	12.5
6	360×288	160×160	6.25
7	180×144	80×80	25
8	180×144	80×80	12.5
9	180×144	80×80	6.25

Table 2.3: Set of cameras used for testing the CRDS.

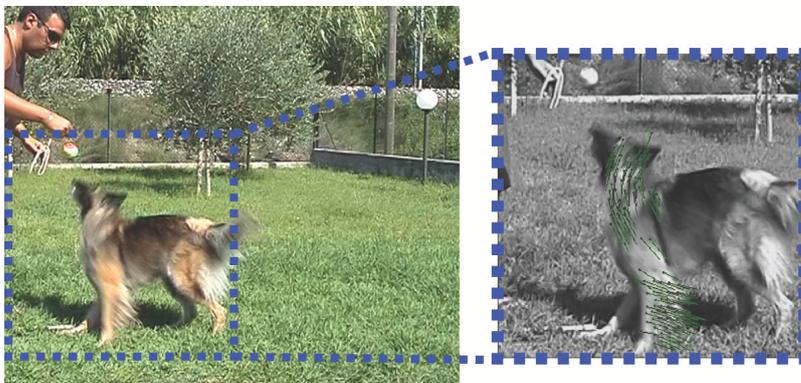


Figure 2.7: Left: Frame extracted from  $C_1$ . Right: the related AoI.

## 2.2.5 Experimental results

### VSN Parameters And AoI Description

The CRDS algorithm has been experimentally validated on the videos of nine different cameras. In Tab. 2.3 the camera frame rate, the image dimensions, and the related AoI dimensions are summarized. The cameras are placed with the same pose respect to the object of interest, which is placed to a distance  $d' = 5$  m. For simplicity, the video from a high resolution and frame rate camera,  $C_1$ , is firstly acquired. Then, the video camera streams  $C_2, \dots, C_9$  have been virtually obtained by applying frame decimation and image resizing techniques on  $C_1$ . For example, the video sequence provided by  $C_6$  is obtained applying spatial decimation factor equal to 2 and a temporal decimation factor equal to 4 on  $C_1$ . In Fig. 2.7 a frame captured by  $C_1$  and related AoI are provided.

Focal Length $f$ [pixels]	$FoV_h$ [rad]	$FoV_v$ [rad]	$q_{lin}$ [ppm]
1200	0.58	0.47	251

Table 2.4: The  $C_1$  parameters and the linear pixel per meter  $q_{lin}$  at a distance  $d' = 5$  m.

### Object Of Interest Velocity Extraction

In this section, the feature velocity extraction technique is described to define the mean velocity of the object of interest within the AoI. In the first step, camera  $C_1$  is calibrated (see Tab. 2.3) through the Camera Calibration Toolbox for Matlab<sup>®</sup> provided in [31]. Thanks to the calibration process, the camera focal length  $f$  is defined and the FoVs is derived through geometrical evaluation. Then, the number of linear pixel per meter is calculated, as defined in (2.4), at the considered distance  $d' = 5$  m. In Tab. 2.4, the  $C_1$  parameters is listed as well as the visualized number of linear pixel per meter. In the second step, the optical flow of the considered video sequence is calculated. The optical flow describes the apparent motion of image objects between two consecutive frames caused by the object or camera movement. For each pixel sub-block in the frame, a 2D displacement vector field describes the movement of the sub-block from the first frame to the second one. A pixel sub-block  $I(x, y, t)$  located in the first frame is considered that it moves by a distance  $(dx, dy)$  in the next frame in a  $dt$  time interval (clearly,  $dt = Fr^{-1}$ ). Assuming the sub-block pixel intensity does not change, it is:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (2.12)$$

Then, taking in consideration the Taylor series approximation of right-hand side of (2.12) removing common terms and dividing by  $dt$ , the following equation is obtained:

$$f_x u + f_y v + f_t = 0 \quad (2.13)$$

where:

$$\begin{aligned} f_x &= \frac{df}{dx}; & f_y &= \frac{df}{dy} \\ u &= \frac{dx}{dt}; & v &= \frac{dy}{dt} \end{aligned} \quad (2.14)$$

In (2.13), the 2D array  $(u, v)$  describes the displacement of each sub-block from the first frame to the next one. To solve (2.13) several numerical algorithms have been proposed in literature. In this work, the optical flow is computed using the *OpenCV* library [32]. In particular, a technique to find the dense optical flow is

utilized based on Gunner Farneback’s algorithm [33]. The optical flow is computed for each pair of consecutive frames provided by  $C_1$ . Then, for each  $i^{th}$  sub-block belonging to the AoI the related velocity is calculated in the following way:

$$v_i = \|m_i\|^2 \cdot q_{lin}^{-1} \cdot F_r \quad (2.15)$$

where  $\|m_i\|^2$  is the displacement magnitude of the considered sub-block and  $q_{lin} = N_1/M$  is the number of linear pixels used for describing a single meter at the considered distance  $d' = 5$  m. In Fig. 2.7-right the displacement vectors in the AoI are indicated by the bold arrows. Then, to compute the average velocity  $\bar{v}$ , as requested in (2.9), I use the fact that, during a video visualization, the attention of the user is more focused on objects that are moving. In other words, the human-eye is more attracted from rapid object position variations, and less sensitive to what-is-happening in the background and in the contour features. Taking into account this behavior, the average velocity has been expurgated from all the features characterized by low or approximately null quantity of movement [32]. With this approach, the average object of interest velocity for our experimental video sequences resulted to be  $\bar{v} = 5.7$  m/s.

## 2.2.6 Subjective Video Quality Assessment Results

Subjective tests are typically used in the video processing field to obtain the human user’s perception of the quality of the processed video sequences. During a Mean Opinion Score (MOS) one or more viewers see different short video sequences on the screen that is in front of them. Each time a sequence is shown, it is asked the viewers to judge their quality by using a determined level scale [34]. In this work, the video quality assessment will be used as reference metric to rank the cameras based on a perceived quality point of view. A modified MOS version is adopted, namely PMOS, as proposed in [35]. In PMOS, the quality assessment is obtained visualising a pair of video sequences at each time. It is asked the viewer to express his/her preference choosing the video sequence with the highest perceived quality. Each possible pairwise video combination is visualized and judged by the voters. A population of 25 voters, equally divided between males and females, has been selected. The camera set is characterized by different spatial and temporal resolution and this makes a direct quality comparison of the AoI unfeasible. For this reason, the scene of interest is extracted based on the position of the point of interest and the camera position and orientation. Then, the video sequences focusing the scene of interest are re-sampled in the spatial and temporal domain to a common format, so that they can be directly compared through objective and subjective tests. In Tab. 2.5 the PMOS results have been proposed. Specifically, each element  $C_{i,j}$  of Tab. 2.5 indicates the percentage of users which prefer the

$C_{i,j}$	1	2	3	4	5	6	7	8	9
1		60	65	90	100	65	100	65	100
2	40		65	30	90	100	65	65	100
3	35	35		40	15	45	20	30	65
4	10	70	60		100	65	100	65	60
5	0	10	85	0		65	65	100	65
6	35	0	55	35	35		65	80	60
7	0	35	80	0	35	35		80	95
8	35	35	70	35	0	20	20		90
9	0	0	35	40	35	40	50	10	

Table 2.5: PMOS results.

video sequence provided by  $C_i$  over  $C_j$ . For example, the 70% of voters prefers  $C_4$  over  $C_2$ . In the final step, the Kemeny-Young method is applied to obtain a camera ranking based on the PMOS results. The Kemeny-Young technique has been developed to identify the most popular choices in an election exploiting a pairwise comparison and assigning a score to all possible ranking sequences. Each sequence considers which choice might be most popular, which choice might be second most popular, and so down to which choice might be least-popular. The ranking sequence obtaining the highest score is the selected one [36, 37].

### 2.2.7 Camera Ranking Techniques Comparison

The comparison between two different sorting metrics is a non-trivial problem: a great number of works has been proposed in literature with the aim at measuring the difference between sorting techniques. In this work, two known metrics are implemented, such as the Spearman’s rank correlation coefficient and the Kendall’s tau distance. These metrics calculate the possible correlation (or distance) between two different ranking strategies. In statistics, Spearman’s Correlation Coefficient (SCC) is a non-parametric measure of statistical dependence between two variables [38]. The Kendall’s Tau Distance (KTD) measures the total number of pairwise inversion between two ranking lists. The larger the distance, the more dissimilar the two lists are [39]. However, these metrics do not consider the ranking positions as well as the element relevance in the pairwise inversion counting: due to this reason, such metrics are called *invariant*. A general solution of this problem consists of weighting each ranking inversion by taking into account the element positional information or defining an element relevance criteria [40]. Intuitively, a ranking inversion on a high-weight element should be more penalizing than an inversion on a low-weight element. Following this approach, a Weighted Kendall’s Tau

Distance (WKTD) version is defined in which the element relevance is determined by the subjective test results. Let  $[n] = 1, \dots, n$  be a set of elements and  $W_n$  be the set of permutations on  $[n]$ . In its turn,  $\psi \in W_n$  is the permutation provided by the PMOS results and  $\psi(i)$  is the ranking of the  $i^{th}$  elements. Again,  $\sigma \in W_n$  is the permutation provided by one the proposed camera ranking algorithms. The WKTD resulting distance between rankings,  $\bar{K}(\psi, \sigma)$ , is defined as:

$$\bar{K}(\psi, \sigma) = \frac{1}{\xi} \sum_{(i,j):i<j} K_{i,j}(\psi, \sigma) \quad (2.16)$$

where:

$$K_{i,j}(\psi, \sigma) = |w_{i,j}(\psi(i) < \psi(j) \wedge \sigma(i) > \sigma(j)) \vee w_{j,i}(\psi(i) > \psi(j) \wedge \sigma(i) < \sigma(j))|. \quad (2.17)$$

In particular,  $K_{i,j}(\psi, \sigma) = 0$  if element  $i$  and  $j$  are in the same order in both evaluated rankings. Otherwise, a penalty is counted setting  $K_{i,j}(\psi, \sigma) = 1$ . In the classical KTD formulation  $w_{i,j} = 1$  and  $w_{j,i} = 1$  whereas  $\xi = n(n-1)/2$  is a normalization factor which counts all the possible ranking inversions. In Tab. 2.5,  $C_{i,j}$  is the percentage of users who preferred element  $i$  over element  $j$ . Therefore,  $C_{i,j}$  is also intended as the percentage of users who may be unsatisfied if element  $i$  and  $j$  are ranked in the opposite order. Hence, the WKTD is proposed setting the weights  $w_{i,j} = C_{i,j}$  and  $w_{j,i} = C_{j,i}$  in (2.16): the higher is the user preference on a certain ranking order, the more significant is the pairwise inversion counting in the considered metric. Accordingly, the normalization factor  $\xi$  is calculated in (2.18) taking into account all the possible pairwise inversions, as

$$\xi = \sum_{(i,j):i<j} |C_{i,j}(\psi(i) < \psi(j)) \vee C_{j,i}(\psi(i) > \psi(j))|. \quad (2.18)$$

The SCC computes the distance between the two permutations in the following way:

$$\rho(\psi, \sigma) = 1 - \frac{6 \sum_{i=1}^n d(i)^2}{n^2(n^2 - 1)} \quad (2.19)$$

where  $d(i) = \psi(i) - \sigma(i)$  is the difference between the two rankings of the  $i^{th}$  element.

## 2.2.8 Camera Ranking Validation

In Tab. 2.6 the camera ranking results are collected considering the objective metrics, the CRSS and the proposed CRDS. In particular, the position of camera

$C_i$	PMOS	PSNR	SSIM	CRSS	CRDS
1	1	1	1	1	1
2	3	2	2	2	3
3	8	5	4	3	7
4	2	3	3	4	2
5	4	4	5	5	4
6	5	6	6	6	8
7	6	7	7	7	5
8	7	8	8	8	6
9	9	9	9	9	9

Table 2.6: Subjective test results (first column) and camera ranking results for the implemented techniques.

$C_i$  in each considered ranking techniques is provided. Furthermore, the subjective test results are also provided based on the Kemeny-Young method in the first column of Tab. 2.6. It is noticeable how the CRSS results mainly depend on the camera resolutions. This is due to the fact that the camera set is characterized by the same pose and this algorithm does not depend on the frame rate and the feature velocities. In its turn, several inversions occur if the ranking provided by the subjective tests are compared with the objective ones. In particular, we can observe how the objective metrics tend to favour high-resolution cameras; conversely, the subjective tests privilege cameras characterized by higher frame rate. The CRDS algorithm validation is composed by two steps. In the first step, we verified the average feature velocity computation and the feasibility of the assumption we made in Sec. 2.2.5. To this purpose, several CRDS results are provided varying the average feature velocity  $\bar{v}$  in (2.9). In Fig. 2.8 these results are compared with the subjective tests using the KTD, WKTD and the complementary version of the Spearman’s coefficient (C-SCC). We can notice that the application of CRDS using the average expurgated velocity we calculated in Sec. 2.2.5 provides the best similarity between CRDS and subjective test results. In the second step, we provided a detailed evaluation of the proposed CRDS comparing the obtained results with the users’ perceived video quality defined by the subjective tests. We can observe how the CRDS and the subjective test ranking results are identical for the top four cameras (*i.e.*,  $C_{1,2,4,5}$ ) which can be identified as the best quality cameras. The main difference is related to the low-quality cameras which are  $C_{6,7,8}$ . Coherently, the proposed ranking algorithm tends to prefer higher frame rate cameras again, while the users’ preference privilege lower frame rate cameras but higher resolution cameras in this case. In Tab. 2.7, the KTD, WKTD and SCC values are presented comparing the ranking based on subjective tests and the other proposed ranking techniques. We can notice that CRDS provides the best results, *i.e.*, the lowest distances and the highest correlation, when compared with the human ranking.

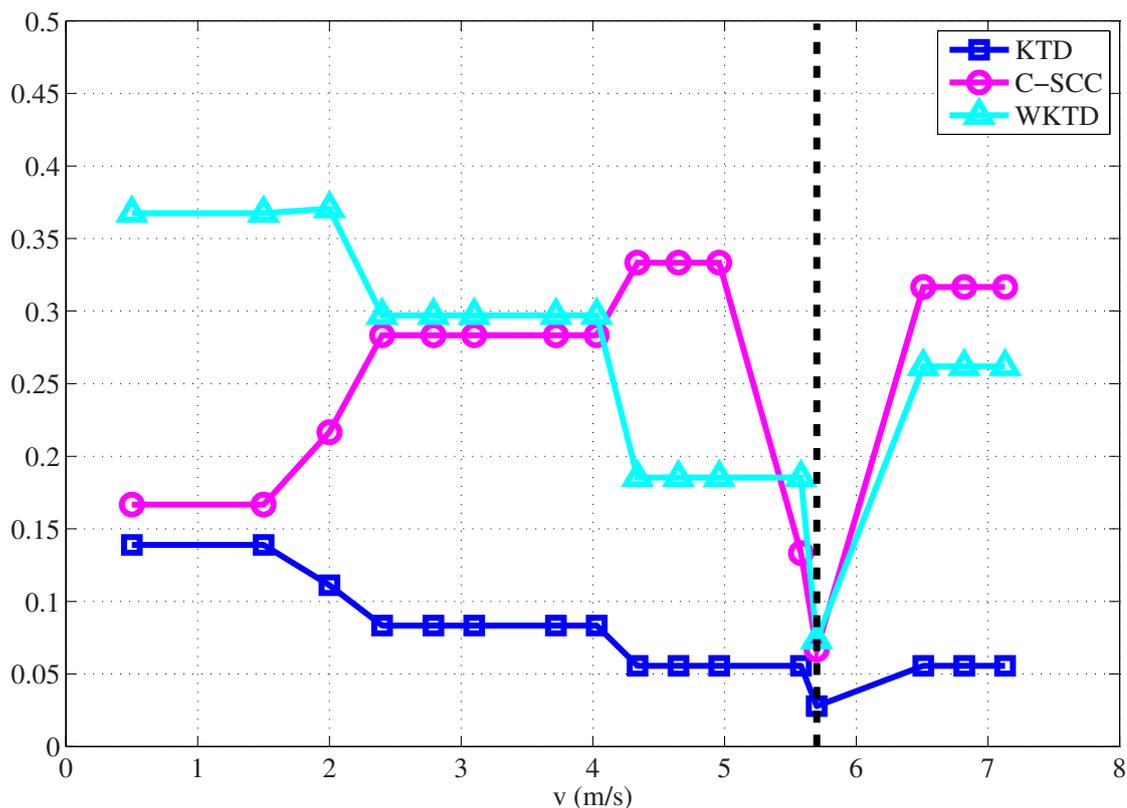


Figure 2.8: Distances (KTD, C-SCC, and WKTD) between CRDS and the subjective tests, as a function of the parameter  $\bar{v}$  in (2.9), for the recorded scene.

## 2.2.9 Camera Ranking Evaluation Tools

In this section the developed Matlab<sup>®</sup>simulation toolbox is described in order to test the proposed camera ranking algorithms. In Fig. 2.9-A the main framework interface has been depicted and the principal controls have been highlighted. In particular, section A is designed to set the size of the monitored area and its position with respect to a given system reference. In Fig. 2.9-B, the area to be monitored is represented by a blue parallelepiped. Furthermore, the monitored area is discretized in sub-volumes: the proposed algorithms work on the single sub-volumes, selecting the video camera which best fits the camera ranking requirements for each of them. The number of sub-volumes can be determined setting the number of discretization points, per dimension, using the section A of the main simulator interface. The larger is the number of discrete points, the more accurate is the camera ranking. On the contrary, a limited number of points entails a lower computational time. In addition, in Fig. 2.9-B the cameras constituting the network have been depicted: in particular, the cameras orientation is highlighted

	PSNR	SSIM	CRSS	CRDS
KTD	0.05	0.11	0.14	0.02
SCC	0.78	0.78	0.84	0.93
WKTD	0.09	0.14	0.19	0.02

Table 2.7: Distances and correlation between automatic and human ranking with  $\bar{v} = 5.7$  m/s.

Fig. A

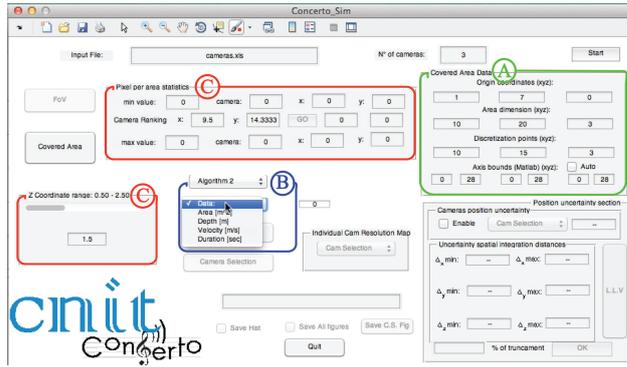


Fig. B

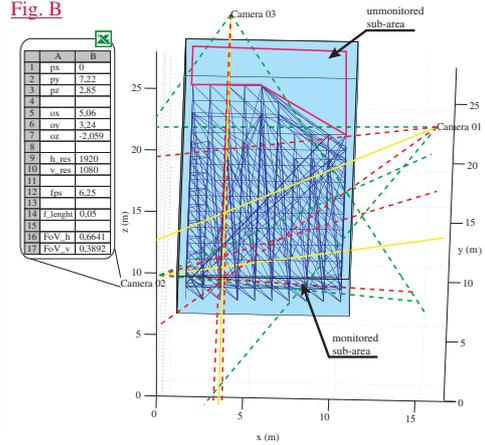


Figure 2.9: The multi camera simulator main interface.

by yellow lines while the horizontal and vertical FoVs are depicted in dashed green and red lines, respectively. Defining the FoVs in association with the camera orientation permits to calculate which portion of the area of interest is monitored by a given camera. Taking advantage of the previous consideration, the ranking algorithm performs a preliminary camera selection discarding all the area sub-volumes not monitored by the camera network. For example, in Fig. 2.9-B the sub-area, depicted in dark-blue, is monitored by the camera network while the upper area results uncovered. All the information relative to each camera is collected in a description file text. In the initial phase, the simulator reads the file associated to each camera in order to deploy the video sensors on the right position. To simulate a realistic VSN, each camera is placed following the pinhole camera model as illustrated in Fig. 2.2. In particular the camera information file contains the following data:

- camera position:  $\mathbf{p} = [p_x, p_y, p_z]$ ,
- camera orientation:  $\mathbf{u} = [u_x, u_y, u_z]$ ,
- horizontal image plane dimension (in pixel):  $h_{res}$ ,

- vertical image plane dimension (in pixel):  $v_{res}$ ,
- frame rate:  $F_r$ ,
- focal length:  $f$ ,
- vertical FoV:  $FoV_v$ ,
- vertical FoV:  $FoV_h$ .

Once the monitored area is set and the camera network is deployed, the two proposed ranking algorithms can be selected in section B, as highlighted in Fig. 2.9-A. In addition, according to the selected algorithm, additional parameters can be set before running the camera selection [11]. Finally, the camera selection results

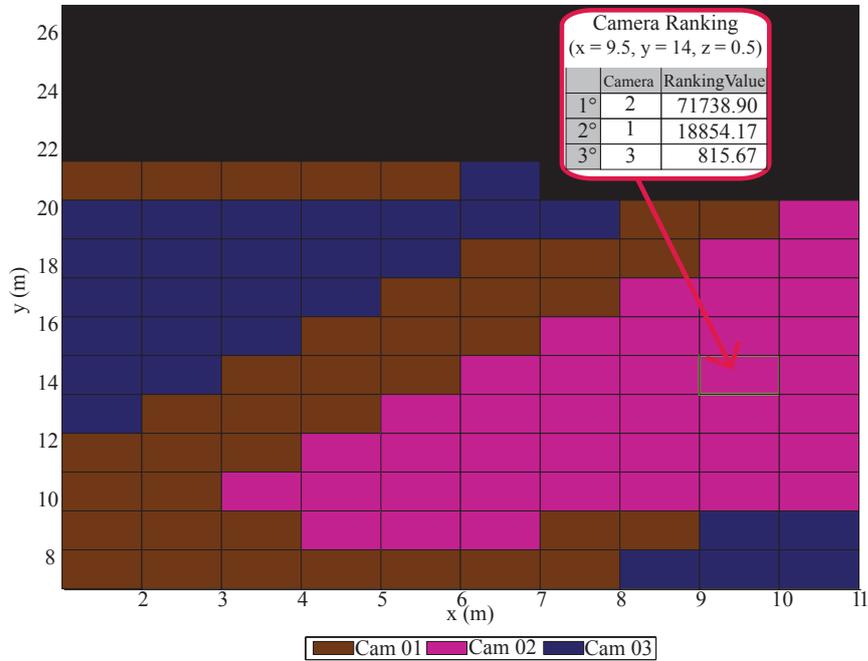


Figure 2.10: An example of camera ranking result.

can be visualized as illustrated in Fig. 2.10. In particular, the figure shows the camera selection for a set of sub-volumes lying on the same plane perpendicular to the z-axis. The correspondent plane can be selected in section C of the main simulator interface. Furthermore, using section C in Fig. 2.9-A, camera ranking results can be displayed for a single sub-volume. For example, in Fig. 2.10 all the cameras are listed with the corresponding ranking value.

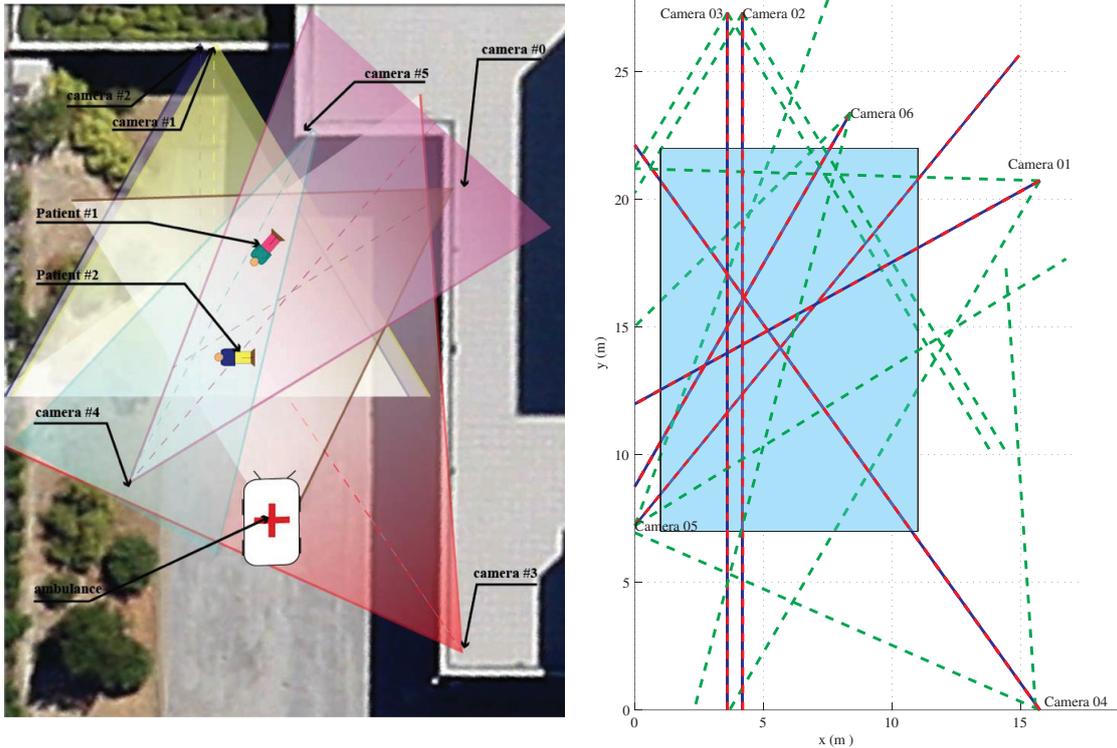


Figure 2.11: Left: the reproduction of the emergency scenario. Right: the simulated scenario version.

## Camera Ranking Experimental Evaluation

The results here reported are obtained reproducing the realistic emergency area at the Hospital of Perugia. In this emergency scenario, a multi-camera network has been deployed in order to monitor the preliminary first-aid operations on a couple of injured people lying on the ground. The camera ranking algorithms are tested on the acquired video sequences, selecting the camera capable to provide the best video quality. In Fig. 2.11 the emergency scenario has been depicted; additionally the camera displacement is illustrated and the position of the ambulance and the injured people. At the right side of the figure, the scenario is reproduced by the Matlab<sup>®</sup> simulator presented in (2.2.9). Moreover, in Tab. 2.8 some parameters of the involved cameras in the simulation are collected. In Fig. 2.12 the scene of interest captured by the different cameras is showed, while in Fig. 2.14 the camera ranking map is depicted, in which the best selected camera capable to cover each different sub-volume are illustrated. In particular, in Tab. 2.9 the camera ranking obtained by the two algorithms for the sub-volume highlighted in Fig. 2.12 are presented. CRSS selects Camera 4 although other cameras are closer to the

Camera Nr.	Resolution	$FoV_h$ [rad]	$FoV_h$ [rad]	frame-per-second (fps)
0	1024x720	1.074	0.72	25
1	1024x720	1.074	0.72	25
2	1024x720	1.074	0.72	25
3	1024x720	1.074	0.72	25
4	1920x1080	0.6641	0.3892	6.25
5	720x576	0.5318	0.43.9	25

Table 2.8: Some parameters of the camera involved in the emergency scenario.

Ranking Position	CRSS	CRDS
1	Camera 4	Camera 5
2	Camera 5	Camera 4
3	Camera 1	Camera 1
4	Camera 2	Camera 2
5	Camera 0	Camera 0
6	Camera 3	Camera 3

Table 2.9: The camera ranking algorithms results.

considered target. Given its higher resolution, Camera 4 guarantees the best trade-off between distance and resolution providing the best quality in visualization. On the contrary, CRDS ranked at the first position Camera 5 although it has the lower resolution with respect to Camera 4. The ranking inversion between Camera 5 and Camera 4 depends on the fact that CRDS takes into account also the video frame rate and the amount of motion within the scene of interest. In fact, Camera 4 has a full HD resolution (1920x1080, very good spatial information), but just 6.25 fps (poor temporal information), while camera 5 has still a good spatial resolution (720x576) but a higher frame rate (25 fps). Finally, two frames extracted from the videos selected by the two algorithms, are showed in Fig. 2.13.

### Key Performance Metrics

The proposed multi-camera strategies are being integrated in the CONCERTO simulator, specifically in the Coordination Center (see Sec. 2.2.3 and in particular Fig. 1.2). To evaluate the effectiveness of the proposed algorithms, different sets of simulations using cameras that are characterized by different spatial and temporal resolution are performed. At the end of each simulation, the proposed camera



Figure 2.12: The scene of interest captured by the different cameras.

ranking algorithms will be validated through objective metrics (PSNR and SSIM) and subjective tests, such as MOS and PMOS. Different cameras are characterized by different spatial and temporal resolution and located at different distance from the point of interest, and this makes a direct quality comparison unfeasible. For this reason, the scene of interest is extracted based on the position of the point of interest and the camera position and orientation. Then, the video sequences focusing the scene of interest are re-sampled in the spatial and temporal domain to a common format, so that they can be directly compared through objective and subjective tests in Fig. 2.15. To this purpose, a C++ library for the video editing of the acquired sequences has been developed: this software tool comprises video up-sampling and down-sampling for both the spatial and the temporal domain. In Fig. 2.16, the procedure that is applied to evaluate the multi-camera ranking performance is showed based on the outcomes of the CONCERTO simulator (see Ch. 4). In particular, the implemented video editing techniques are:

- pixel repetition for the up-sampling (Fig. 2.16-b), pixel decimations for the down-sampling (Fig. 2.16-c),
- pixel interpolation with ideal-sync kernel for both up-sampling and down-

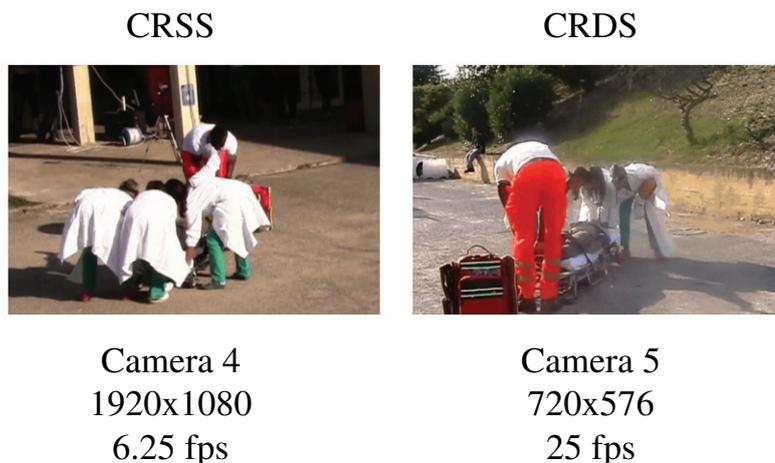


Figure 2.13: The selected cameras by the proposed algorithms.

sampling (Fig. 2.16-d and Fig. 2.16-e, respectively ),

- up-sampling with Lanczos kernel (Fig. 2.16-f), down-sampling with the Inter-Area technique which re-sampling using pixel area relationship (Fig. 2.16-g).

For the temporal domain, the implemented techniques to modify the video frame-rate are:

- frame repetition for the up-sampling, frame decimations for the down-sampling,
- frame interpolation with Lanczos kernel,
- frame interpolation with linear kernel.

The performance achieved considering the full architecture are presented in Sec. 2.2.2 and in Sec. 2.2.3 are evaluated in Ch. 4, by implementing the ranking techniques in the CC of the CONCERTO simulator and will be reported in [12].

## 2.3 Conclusions

In this Chapter, two novel camera ranking algorithms have been presented. The proposed algorithms have been developed with the aim to find, within a VSN, the camera that best satisfies specific ranking criteria. The proposed algorithms are designed to optimize the QoE for the final user in terms of video quality and significance. To test the camera selection techniques, an indoor multi-camera acquisition system is arranged and used the resulting video sequences in order to

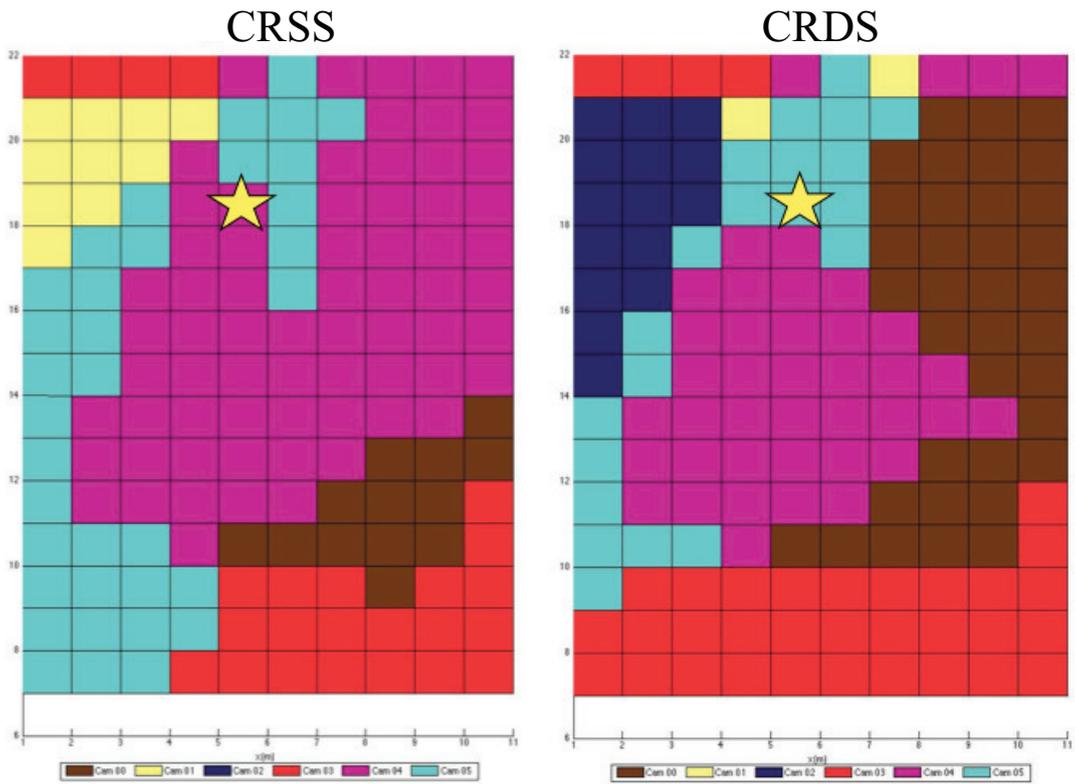


Figure 2.14: The camera ranking algorithms results for the whole emergency area.

perform some evaluations on the camera ranking algorithms. Numerical results are derived by experimental simulations and subjective tests which testify the efficiency of the proposed techniques in providing the best final user experience in terms of visual quality in contexts of healthcare and security application scenarios.

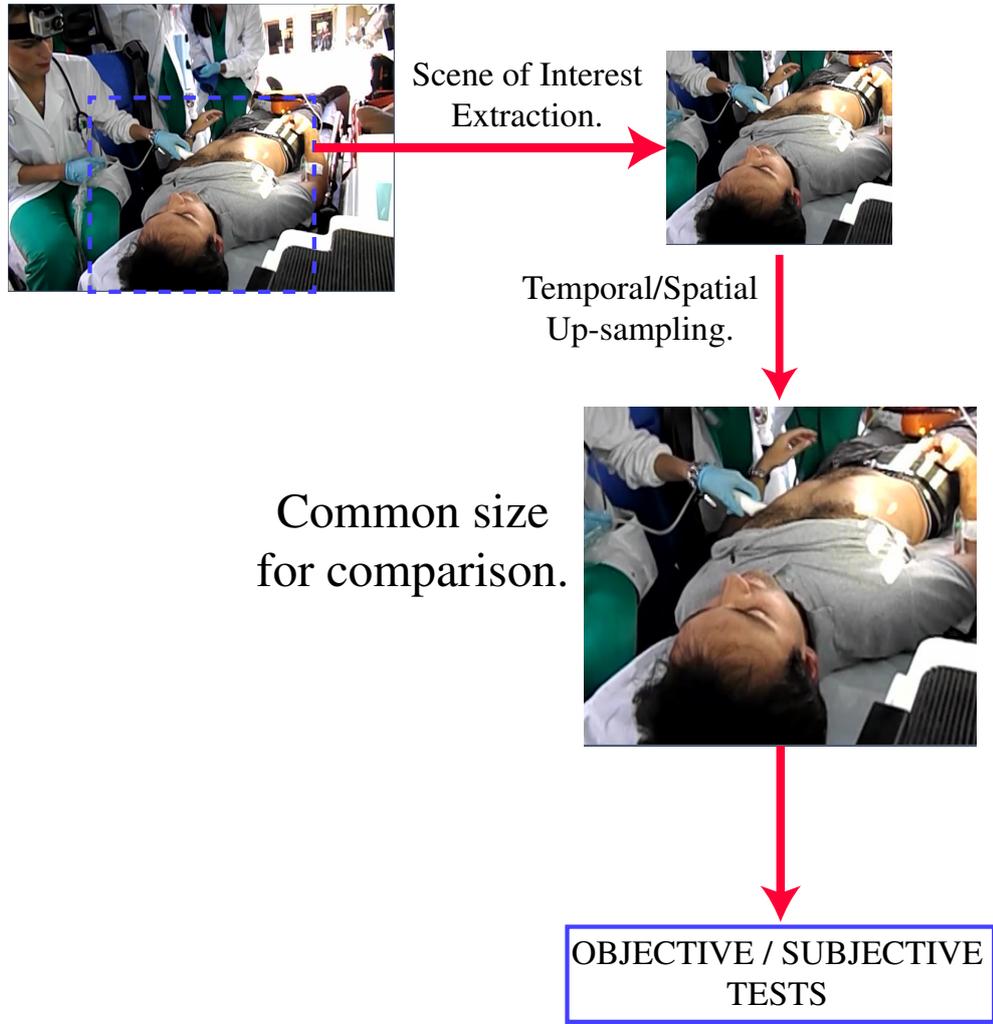


Figure 2.15: The re-sampling technique suitable for a fair comparison between different Scene of Interest cameras.



(a) Original Frame



(b) Pixel Decimations



(c) Pixel Repetition



(d) Sync down-sampling interpolation



(e) Sync up-sampling interpolation



(f) Inter-area spatial down-sampling



(g) Lanczos up-sampling interpolation

Figure 2.16: The different implemented techniques for the spatial video up/down-sampling.

# Chapter 3

## Video Adaptation And Transmission Techniques For The support Of mHealth Services

### 3.1 Introduction And State-of-the-art

The emerging concept of Mobile Health (mHealth) represents a natural evolution of Electronic Health (eHealth) from the conventional tele-medicine to the progressing applications of mobile and wireless telecommunication. Being part of modern tele-medicine which generally offers higher diagnosis and treatment quality standards, reduces medical costs, and provides possibilities to handle problems of the aging human society – mHealth rises in parallel with the rapid adoption of mobile communications and computing, sensor networking, and advanced wireless technologies into our daily life. Although mHealth is explained as a subgroup of eHealth, it is commonly considered as a unique healthcare paradigm thanks to its special capabilities to provide ubiquitous connectivity, personalized and accurate provision of services, automated and expedited healthcare delivery and reduced costs, such offering patients with more practicality and convenience [41]. Recent years have witnessed an increasing and remarkable advancement of mHealth systems and applications, mainly driven by the fact that for a significant part of the world's population access to various healthcare services is commonly limited by geographical distance, cost, and availability of qualified medical personnel, and by other challenges of societies like declining work force due to the ageing population, fast development of global medical knowledge coming with expensive new diagnostic procedures and infrastructure, and ever-growing expectations for the highest quality healthcare. New, immersive, multimedia-driven, pervasive and interactive mHealth services are rising to provide timely and prompt medical attention while

also saving monetary resources. The key areas explored are how mHealth can facilitate wearable sensor system based personal healthcare monitoring, self-diagnosis, early detection and preventive care [42]; mobile-assisted tele-rehabilitation and therapy [43]; monitoring of soldiers in tactical environments [44]; medical care in emergency and mass casualty situations [45]; intelligent home care [46]; mobile robotic systems for controlling medical devices in isolated areas [47]; and other specialized services like drug intake reminder, raising awareness of health issues, performing health surveys, maintenance of personalized medical records, and medical advice provision/tele-consultation [13].

However, like other services running in mobile and wireless environments, the efficiency and usability of mHealth applications are substantially impacted by the continuously varying environmental characteristics, scarce network resources, sparse radio bands and bandwidth, battery and computational power of mobile devices, fluctuating delay, jitter and other QoS parameter values. This clearly implies the need of scalable, context and content-aware mechanisms making applications, service provision and delivery procedures adaptable to the extremely diverse mobile environments. The impacts of fluctuations in context and scarcity of resources are further aggravated by the current trends that prognosticate a massive traffic volume growth in mobile telecommunication systems during 2011-2020 [48]. To date, this traffic explosion is mostly driven by Internet applications providing almost an unlimited scale of interaction, information, and entertainment services for human users. But with the widespread deployment of autonomous, networked and inter-operating sensor technologies, another form of communications called Machine-to-Machine (M2M) or Machine Type Communication (MTC) is emerging nowadays which supposedly will be the leading traffic contributor for mobile Internet evolution in the near future and also has the potential to become a major enabler of successful live mHealth deployments [49]. Hundreds of billions of intelligent, resource constrained sensors and other devices are envisioned to be interacting without human intervention in the Internet of Things (IoT), and generating an enormous amount of data for monitoring, measuring, biomedical signal processing, remote controlling/intervention, safety, or surveillance functions in mHealth or other advanced application areas. An other prominent driving force of mobile traffic evolution is the advancement of high bit-rate multimedia applications. It is foreseen that due to the expansion of data-hungry services like television/radio broadcasting and high-definition Video on Demand, mobile video traffic volume will increase 25 fold between 2011-2016, accounting for over 70 percent of total mobile data traffic by the middle of the decade [48]. This trend is also visible in mHealth: the spreading of multimedia technologies and developments in mobile connectivity gives doctors, hospitals and medical institutions a new set of tools for managing patient care, disease history and billing of medical services, using 2D/3D

imaging (*e.g.*, X-ray and ultrasound scans, computer tomographs and magnetic resonance images) for diagnostic purposes and providing radically new types of multimedia services, such as multi-view, stereoscopic and holographic video communications for tele-diagnosis or even helping remote participation of prominent experts in life-saving operations [13].

The scale of the traffic volume and IoT expansion together with the extremely high requirements of diagnostically useful multimedia transmission techniques poses serious challenges for mobility management of mHealth. The principal problems, which are necessary to solve for mobile health to be a common medical and public health practice, are as follows: 1) provide mobile users with ubiquitous communication capabilities over any heterogeneous access infrastructures; 2) offer medical QoS/QoE for multimedia data transmission during continuously varying network conditions; 3) deliver real-time medical data reliably, even in vertical (*i.e.*, inter-system) handover situations; 4) optimize overall system performance within a multilayer/cross-layer management infrastructure [13]. The remainder of this chapter is organized as follows: in Sec. 3.1.1 and in Sec. 3.1.2 an overview of the state-of-the-art is provided concerning the radio resource allocation and scheduling techniques in LTE Uplink environment and the cross-layer video adaptation techniques respectively. In addition, in Sec. 3.2 an SVC cross-layer optimization solution is provided for mHealth scenarios that can be implemented in the use case described in Sec. 1.1. Moreover, in Sec. 3.3 a joint combination between the camera ranking algorithms is proposed in Ch. 2 and the video adaptation schemes in this Chapter are provided with the aim at satisfying the quality constraints and maximizing the QoE of the use case described in Sec. 1.2. Furthermore, the work in Sec. 3.3 is extended in Sec. 3.4 considering a more complex camera ranking scenarios in which the proposed solutions are dynamically applied taking into account multiple points of interest.

### **3.1.1 Radio Resource Allocation And Scheduling In LTE Uplink**

The preliminary definition of the scenario use cases in Sec. 1.1 and in Sec. 1.2 provides different sets of QoS requirements, in terms of guaranteed bit-rate, maximum end-to-end delay and packet loss. Among them, the studied scenarios should exploit the uplink of existing and next-generation wireless systems. While the Third Generation (3G) cellular technologies only offer QoS support for voice services, beyond-3G and 4G systems, such as the LTE [50] and its upgraded versions, are expected to provide QoS support for multimedia application according to different Class Quality Indicators (CQI). In particular, the Technical Specification (TS) 23.203 [51] provides different class-of-service priorities in terms of minimum

GBR, Packet Delay Budget (PDB) and Packet Error Loss Rate (PELR). According to the proposed requirements for above-mentioned use cases, mobile terminals should receive a priority value at least equal to 4, corresponding to packet delay and packet error rate lower than 150 ms and  $10^{-3}$ , respectively, with a minimum guaranteed bit-rate. However, none of such requirements can be accomplished without efficient Radio Resource Allocation (RRA) and multiuser scheduling at the MAC layer for both the downlink and the uplink, and these matters are not defined by the LTE standards. Assuming a cross-layer exchange of information, such as as Channel State Information (CSI) from the PHY layer and/or source content information from the upper layers, RRA strategies are often categorized in channel-aware and/or content/context-aware schedulers. A channel-aware scheduler takes advantage of the temporal, frequency and multi-user diversity inherently provided by the wireless channel. Each user in a cellular scenario experiences different channel gains on the available radio resources, according to path loss, shadowing and fast fading. Therefore the scheduler can exploit such information to allocate the radio resources to those users whose channel conditions are favourable. The content-aware schedulers take into account the specific content (or context) of the source data, such as, for example, rate-distortion relationship and play-out deadline of video sources, in order to provide differentiated services for heterogeneous applications. While the research community has provided a large amount of proposals to solve the RRA problem for LTE/LTE-like downlink, the uplink RRA problem has gathered attention only starting from 2004 and a sensible lower number of contributions can be found. In the following an overview of the last improvements in the RRA methodologies for the LTE uplink is provided.

## **Radio Resource Techniques For LTE Uplink Overview**

The LTE standard [50] was developed by the 3GPP, starting from 2004, to address the increasing demand for better QoS and the growth of bandwidth consuming multimedia applications. Compared to its predecessors LTE provides several enhancements, as higher peak data-rate and spectral efficiency, flexible resource allocation, lower user and control plane latency and reduced devices complexity. In order to meet such requirements, Orthogonal Frequency Division Multiplexing (OFDM) has been chosen for the first time as the access technology for both downlink and uplink in cellular networks. The OFDM modulation scheme divides the input data stream into several parallel sub-streams with reduced data rate, hence with increased symbol duration, and each sub-stream is transmitted on a separate orthogonal sub-carrier. The higher symbol duration improves the robustness against the channel delay spread due to multi-path propagation. In an OFDM system, transmission resources are available in both time and frequency domain,

in terms of OFDM symbols sub-carriers respectively. The time and frequency resources can be organized into sub-channels for allocation to individual users leading to Orthogonal Frequency Division Multiple Access (OFDMA), also known as multi-user OFDM. Within LTE, uplink transmission is organized into radio frames with 10 ms duration with two different radio frame structures according to the supported duplexing schemes, namely Time Division Duplexing (TDD) and Frequency Division Duplexing (FDD). Each 10 ms radio frame is divided into ten equally sized sub-frames. In its turn, each sub-frame consists of two equally sized slots. The entire bandwidth is divided into several sub-carriers with sub-carrier spacing of 15 kHz, allowing six or seven OFDM symbols for each sub-carrier according to the two supported Cyclic Prefix (CP) durations. In the uplink, three constellations are allowed on each sub-carrier, *i.e.*, Quadrature Phase Shift Keying (QPSK), 16-QAM and 64-QAM. In order to reduce the feedback information and the RRA complexity, adjacent sub-carriers are grouped into PRB of 12 sub-carriers each, which represents the minimum amount of allocable resources [13]. An important drawback of the OFDMA schemes is that the transmitted time-domain waveform exhibits very pronounced envelope fluctuations, resulting in a high Peak-to-average Power Ratio (PAPR). Signals with high PAPR require highly linear power amplifiers to avoid excessive inter-modulation distortion. To achieve this linearity, the amplifiers have to operate with a large back-off from their peak power, resulting in increasing costs and power consumption. Such problem is clearly more critical in the uplink transmission where the cost and power consumption of mobile must be kept as lower as possible. For this reasons, SC-FDMA [52], also known as Discrete Fourier Transform (DFT)-spread Frequency Division Multiple Access (FDMA), has been introduced for the LTE uplink. SC-FDMA offers advantages similar to OFDMA but provides a lower PAPR by introducing a DFT pre-coding process at the transmitter, aimed at spreading the power over the entire allocated bandwidth. Such an advantage is paid with an increased Inter-symbol Interference (ISI) at the receiver, so that some kind of adaptive frequency domain equalization is required. This trade-off is well balanced in 4G cellular systems, since higher costs due to complex signal processing at the eNodeB are acceptable if they entail simple and low-consumption amplifiers in the mobile terminals. There are two types of SC-FDMA: localized-FDMA (L-FDMA), in which the sub-channels assigned to a user are adjacent to each other, and Interleaved-FDMA (I-FDMA) in which users are assigned with sub-channels distributed over the entire frequency band. Only L-FDMA is taken into account by LTE and this imposes an additional constraint (*i.e.*, contiguous sub-carriers) with respect to traditional OFDMA resource allocation. A detailed overview of SC-FDMA schemes can be found in [53, 52].

### 3.1.2 Cross-layer Optimization For Video Streams

The proposed scenarios in Ch. 1 utilize the rate adaptation capability of video streaming to adjust the video bit-rate to network capacity for optimal QoE and network resource usage. This can greatly expand the video service availability especially in wireless networks and mobile context where the used network connection cannot always support the QoS requirements of the whole video stream. In some situations, the ability to communicate with even the lowest quality video may be critical for some of the medical services. Moreover, the rate and congestion adaptation capability of the medical video streaming services is an important tool to avoid causing a congestion collapse in resource limited networks with potentially large amount of concurrent users [13].

#### Overview of Video Adaptation Approaches

For video streaming, the inability of wireless networks to guarantee the required bandwidth and QoS for the services has boosted the development of novel video coding and adaptation solutions to improve the robustness and QoS for video transmission. For instance, the SVC technology [54] provides both bit-rate and device capability adaptation, which are especially useful in heterogeneous network environments. In addition, several algorithms and protocols for controlling the video stream bit-rate to match the available network capacity have been proposed in the literature. The typical solution of adapting video streams in the application layer has been studied, for instance, in [55, 56]. In this case, video bit-stream adaptation takes place in the server or an intermediate network node and the decision-making relies on client feedback information of the streaming performance (*e.g.*, delay and loss metrics). However, due to this very feedback signaling requirement, application layer adaptation is not responsive enough to quick wireless link capacity fluctuations. To overcome the deficiencies of application layer adaptation solutions in wireless networks, several proposals for prioritizing and adapting video streams in the data link layer have appeared during the recent years. Video traffic can be prioritized over other less critical traffic with different MAC layer QoS solutions supported by most wireless standards (*e.g.*, IEEE 802.11 Wireless Local Area Network (WLAN), IEEE 802.16 WiMAX, 3GPP). The standard QoS architectures, however, provide QoS support in per-flow basis, that is, all video packets receive equal QoS treatment from the MAC. This is insufficient for video streams that would benefit from per-packet QoS differentiation in order to ensure that the most important video packets (*e.g.*, SVC base layer ones) get transmitted also under limited transmission capacity. This problem has been addressed in the literature by several proposals for adapting video streams in the MAC layer, including [57, 58, 59, 60, 61]. In the proposals, video adaptation in the MAC

layer employs selective packet discarding and prioritized transmission to ensure that the most important video packets are transmitted over the wireless link with the highest probability. Such capability typically assumes cross-layer signaling between the application and MAC layers to communicate the video packet priorities. However, the proposed solutions do not take into account fairness towards other traffic types. One solution for implementing QoS differentiation within the video traffic category without hampering the QoS for other traffic types in WLAN is provided in the recently finalized IEEE 802.11aa standard [62]. Nevertheless, all the MAC-level solutions regulate video streaming only in the scope of the wireless link, thus potentially wasting transmission resources in the wired core network. Thus, it can be acknowledged that local adaptation within a single system layer is not the most efficient way to achieve dynamic scalability. Therefore, cross-layer video adaptation solutions, as suggested in [13], are employed.

In a cross-layer scheme, both application layer and MAC layer adaptations are implemented, as proposed in [63]. The application layer adaptation is performed by exploiting information on bandwidth availability from lower layers (*e.g.*, using [55]), whereas MAC layer adaptation is done by using side information on video packets, such as SVC layer information, priority and achievable rate-distortion improvement. Regarding the adaptation of multiple video streams, cross-layer optimization has been investigated, within the framework of SVC, in several contributions, *e.g.*, [64]. Most of the proposed optimization schemes aim at minimizing the aggregate distortion of the multiple video programs. However, different video sequences have different complexities, hence the relationship between rate and quality differs from one video sequence to another. Assuming that the physical resources are shared among different video sequences, strategies aimed at minimizing the aggregate distortion of each video could lead to a high level of distortion for the most complex videos, typically requiring higher rates. Nevertheless, the end-user expectation is to receive the best feasible quality independently of the particular video complexity. In order to optimize the transmission strategy based on the end user video quality, the rate should be allocated among the videos based on a fairness criterion [65]. Moreover, the trade-off between the goal of reducing the bit rate and the goal of keeping the distortion at acceptable levels can be afforded dynamically, in order to perform adaptation to different conditions. Some contributions exist in the literature that consider fairness-oriented rate adaptation, but they exploit the Fine Granularity Scalability (FGS) tool, *e.g.*, [66] [67]. However, FGS mode has been removed from SVC, due to its complexity. A multi-stream rate-adaptation scheme aimed at minimizing the distortion at the end-user while preserving fairness is proposed in [65]. It is based on the knowledge of the total bandwidth to be shared by the multiple videos and on an accurate Rate-distortion (R-D) model of medium grain scalable Medium Grain Scalabil-

ity (MGS) encoded video. R-D models enable to predict the minimum bit rate required to achieve a target quality. The rate of a video sequence is expressed in bytes/s, while the distortion is defined in terms of Mean Square Error (MSE). The PSNR is more often used to express the quality of a video sequence. The time required to model the R-D curve for a given sequence may drive the decision on the methodology/algorithm to be adopted for the R-D modeling. On the other hand, the performance of the streaming system is directly affected by the accuracy of the R-D model [68]. R-D models are often categorized in analytical, semi-analytical and empirical models. Analytical R-D models are used to predict rate and distortion of video sequences prior to the encoding process but they often incur in a loss of accuracy. Empirical models require the computation of all R-D points set resulting in a high complexity. Semi-analytical models aim at reducing such complexity by deriving parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve fitting from a subset of the rate-distortion empirical data points. Many R-D models have been proposed in the literature for real time and non-real time video streaming, see for example, [68][64] and references therein. In [69] the authors present a detailed analysis of the R-D relationship in FGS coders and provide an accurate semi-analytical square root R-D model, which requires at least two empirical points. Enhanced analytical R-D models for H.264/Advanced Video Coding (AVC) were proposed for coded video sequences in [64]. However, the parameter extraction is performed after transformation and quantization in the encoding process and the numerical results show that they exhibit a low accuracy. In [70] a technique to further reduce the complexity of semi-analytical models by keeping similar performance on the accuracy has been investigated. The proposed framework introduces new functions dependent only on the uncoded video streams. The coefficients of these new functions can be estimated off-line through a prior knowledge of the parameters of a set of sample video sequences, and then used for any future video sequence. The R-D model proposed in [70] only uses two parameters which are calculated taking into account the characteristics of the video sequences through a spatial and a temporal index extracted from the original raw video streams. This model has been finally applied as a component of a fairness-oriented multi-stream rate adaptation scheme.

## 3.2 Cross-layer Optimization For mHealth SVC Multiple Video Transmission over LTE Uplink

Mhealth services are expected to become increasingly relevant in the management of emergency situations, enabling real-time support of remote medical experts. In this context, the transmission of health-related information from an ambulance to a remote hospital is a challenging task, due to the variability and the limitations of a mobile radio link. In particular, the transmission of multiple video streams can improve the efficacy of the tele-consultation service, but requires a large bandwidth to meet the desired quality, not always guaranteed by the mobile network. This scenario is included in the use case presented in Sec. 1.1. In this regard, a novel cross-layer adaptation strategy is proposed for multiple SVC videos delivered over a single LTE channel, which adapts the overall transmitted throughput to the actual available bandwidth while being able to provide high quality to diagnostic video sequences and lower (but fair) quality to less critical ambient videos. The proposed solution is based on joint video adaptation and aggregation performed directly at the application layer. From the LTE eNodeB perspective, only a single communication link characterized by given QoS guarantees needs to be managed between the ambulance and the remote hospital, while additional spectrum efficiency is gained from video multiplexing. In our solution the adaptation is designed to optimize quality and fairness by exploiting the information on the available rate assigned by the LTE eNodeB. After introducing a realistic LTE uplink scenario, it is shown that the proposed strategy permits to achieve a good end-to-end quality even in the presence of rate limitations and fluctuations due to the wireless channel and intense traffic within the LTE cell [3].

### 3.2.1 System Architecture

The mHealth scenario addressed is depicted in Fig. 3.1 and refers to the *Use case 1: Ambulance and emergency area* is described in 1.1 [7]. An ambulance equipped with multiple cameras and medical devices reaches an emergency area, where one or multiple injured persons need immediate medical assistance. Multiple information is sent through the available LTE radio access network to the emergency management center at the hospital, where specialized medical staff can follow the first-aid operations, coordinate the intervention and acquire the health-state information necessary to prearrange the treatment at the hospital. The doctors at the hospital interact in real-time with the ambulance staff, receiving both ambient and ultrasonography videos. The real-time video streams are constituted by  $N$  ambient videos acquired by a set of cameras installed on the ambulance and one

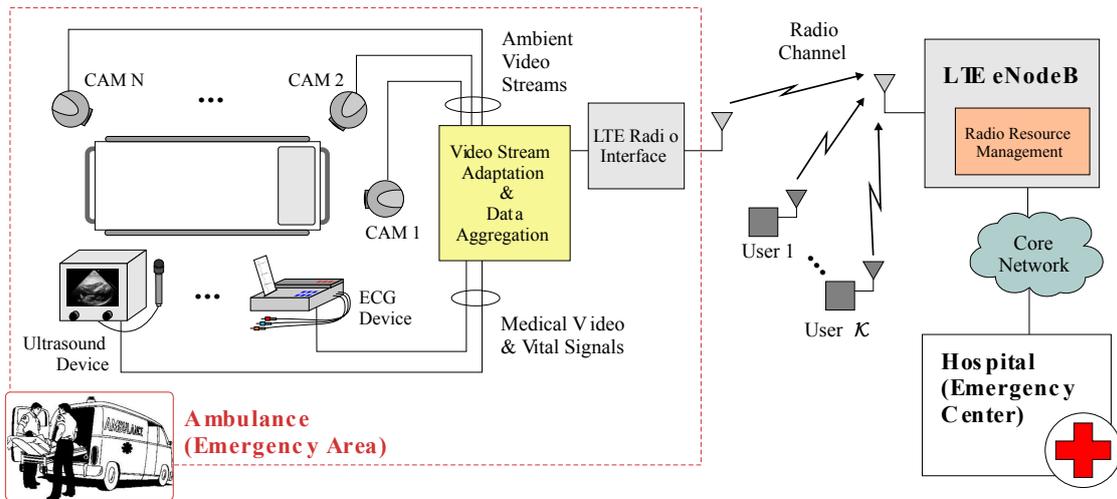


Figure 3.1: The proposed mHealth architecture for emergency scenarios.

diagnostic Focused Assessment with Sonography for Trauma (FAST) sequence. The multimedia flows are processed in real-time and multiplexed with the other medical information by the video adaptation and data aggregation unit before transmission over the LTE radio channel.

From the LTE network perspective, the ambulance (in the following also indicated as m-health user) competes for radio resources with other  $K$  users within the cell, indexed by the set  $\mathcal{K}$ , subdivided into  $K_1$  GBR users and  $K_2$  best-effort users, indexed by the sets  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , respectively. The eNodeB tries to guarantee the transmission rates  $\bar{R}_0$  to the m-health user<sup>1</sup> and  $\bar{R}_k$  to the  $k$ -th GBR user, with  $k \in \mathcal{K}_1$ , while the throughput to best-effort users is provided fairly, according to the resources left after allocating all GBR users.

The video adaptation unit performs two fundamental tasks:

- It manages the inherent different priority of the data flows generated by the m-health user. In particular, it optimally adapts the SVC-encoded streams, in order to deliver the ultrasonography information with sufficiently high quality and the set of ambient videos tuned according to quality fairness criteria.
- It produces an aggregated throughput adapted to the radio channel and cell traffic conditions, according to the amount of resources assigned by the eNodeB to the m-health user.

<sup>1</sup>Note that here as in the rest of this section I indicate the m-health GBR user with the subscript 0.

In order not to complicate the notation, in the following I denote with  $\mathcal{V}$  indifferently the set of  $N + 1$  videos processed by the m-health user and the set of indices identifying these videos.

### 3.2.2 Video Coding And Adaptation

SVC allows the encoding of video sequences into scalable streams. From a scalable stream it is possible to extract a sub-stream in order to meet a given frame rate, resolution and quality, according to three scalability methods supported by SVC. The video sequence is organized in sets of consecutive frames called Groups of Pictures (GOPs). The interval (in frames) between two consequent I-frame is here assumed as multiple of the GOP size.

For the m-health application addressed in this section, the GOP encoding format IBPBPBPP is adopted and exploit temporal and Signal-to-noise Ratio (SNR) scalability with a fixed spatial resolution. More precisely, for high quality ultrasound video temporal scalability is allowed with two available temporal decimations, whereas only SNR scalability is allowed for ambient videos. In SVC, SNR scalability is obtained by enhancing the quality of the video stream with the addition of refinement layers. Two different possibilities are now available in the SVC standard [54], namely the Coarse Grain Scalability (CGS) and the MGS. MGS provides finer granularity with respect to CGS coding; it is obtained by dividing a CGS layer into up to 16 MGS layers. In this approach, I focus on MGS scalability.

Rate adaptation techniques dynamically adapt the amount of transmitted information to the available channel bandwidth by taking into account the content of the videos and its impact on the end user quality. R-D models enable to predict the minimum bit rate required (in bit/s) to achieve a target distortion, defined in terms of MSE. A semi-analytical R-D model is considered: while empirical models require the computation of all R-D point set, resulting in a high complexity, semi-analytical models reduce the complexity by using parametrized analytical functions derived through curve fitting from a subset of the R-D empirical data points or from suitable models connected to temporal and spatial indexes of video sequences, as in [70].

Let us consider the  $v$ -th SNR-scalable video ( $v \in \mathcal{V}$ ) resulting from the encoding of the set of pictures  $\mathcal{I}_v$ . I define  $\mathcal{D}_v = \{d_{1,v}, \dots, d_{E_v,v}\}$  as the set of expected distortion values, one for each extractable sub-stream, whose total number is  $E_v$ . The distortion  $d_{e,v}$ ,  $e = 1, \dots, E_v$ , is computed as MSE between the original and the reconstructed pictures averaged over  $\mathcal{I}_v$ . The minimum information rate in bps,  $F_k(d_{e,v})$ , required to transmit the  $e$ -th sub-stream with the given expected distortion  $d_{e,v}$ , is a function of the discrete values in  $\mathcal{D}_v$ . Following the approach in [71, 64] and recently extended in [65], the expected R-D relationship is modeled by using a parametric function  $F_v(D)$  of a continuous variable  $D$ , defined

over a limited interval, *i.e.*,  $D \in [\bar{D}_v^{\text{bl}}, \bar{D}_v^{\text{hl}}]$ . Within SVC,  $\bar{D}_v^{\text{bl}}$  and  $\bar{D}_v^{\text{hl}}$  are the expected distortions of the set of pictures  $\mathcal{I}_v$  after decoding the base layer (with rate  $\bar{F}_v^{\text{min}}$ ) and the highest enhancement layer (with rate  $\bar{F}_v^{\text{max}}$ ), respectively. Although this framework holds for any monotonic strictly decreasing function  $F_v(D)$ , the following parametric model is considered:

$$F_v(D) = \frac{\alpha_v}{D + \xi_v} + \beta_v, \quad D \in [\bar{D}_v^{\text{bl}}, \bar{D}_v^{\text{hl}}], \quad \xi_v > 0, \quad (3.1)$$

where  $\alpha_v$ ,  $\xi_v$  and  $\beta_v$  are three parameters depending on the temporal and spatial characteristics of the set of pictures  $\mathcal{I}_v$  and on the frame rate.

### 3.2.3 Adaptation Algorithms

It is assumed the ambulance equipment negotiates with the LTE access network a guaranteed bit rate  $\bar{R}_0$  to support the emergency m-health services. Such value  $\bar{R}_0$  is eventually updated every  $W$  seconds in case of critical cell-load conditions or bad channel conditions for the ambulance. The m-health user exploits this guaranteed bandwidth to deliver the best video quality according to priority and fairness constraints among the different videos. This is obtained through a dynamic rate adaptation strategy, consisting in maximizing the overall video quality while minimizing a weighted quality difference among the different videos under minimum and maximum rate constraints. This strategy has been first proposed by the authors in [65] by addressing the problem for both empirical and semi-analytical R-D model and without considering weighted quality difference. In [65] it has been shown that, when the parametric R-D model is sufficiently accurate, it can be used to relax the multi-objective optimization problem leading to a much simpler constraint satisfaction problem. Here, this approach is followed by defining an adaptation strategy based on the solution of the following problem:

$$\sum_{v \in \mathcal{V}} \bar{H} F_v = \bar{R}_0, \quad (3.2a)$$

$$\Delta(D_i, D_j; w_i, w_j) = 0 \quad \forall i, j \in \mathcal{V}, i > j \quad (3.2b)$$

$$F_v^{\text{min}} < F_v < F_v^{\text{max}} \quad \forall v \in \mathcal{V} \quad (3.2c)$$

where  $\mathcal{V}$  is the set of videos handled by the ambulance for e-health emergency services,  $\bar{H}$  takes into account of the estimated overhead introduced at the different layers of the network architecture,  $\Delta(D_i, D_j; w_i, w_j)$  is the distortion-fairness metric for each pair of videos, and  $w_v$ ,  $v \in \mathcal{V}$  are the weights used to account for the different priorities mentioned in the introduction.

The distortion-fairness metric is defined as:

$$\Delta(D_i, D_j; w_i, w_j) = \begin{cases} 0 & \text{if } (i, j) \in \mathcal{X} \vee (j, i) \in \mathcal{X} \\ |w_i D_i - w_j D_j| & \text{otherwise} \end{cases} \quad (3.3)$$

with

$$\mathcal{X} = \{(i, j) \in \mathbb{N}^2 : (D_i = \bar{D}_i^{\text{bl}} \wedge w_i D_j > w_j D_i) \vee (D_i = \bar{D}_i^{\text{hl}} \wedge w_j D_j < w_i D_i)\} \quad (3.4)$$

where  $\wedge$  and  $\vee$  are the "AND" and "OR" operators, respectively. The definition of the fairness metric for a pair of videos  $i$  and  $j$  takes into account that if  $D_i$  has already achieved its minimum value, but the rate of the  $j$ -th video can be further increased to improve the sum-rate, then the fairness metric is set to zero, allowing such additional flexibility. The converse is also possible if  $D_i$  has already achieved its maximum value, but the rate of the  $j$ -th video can be further decreased to make sum-rate achievable.

The algorithm proposed in [65] can be suitable extended to solve the problem considered here. Let  $x_v, y_v \in \{0, 1\}$ ,  $v \in \mathcal{V}$ , be binary variables that indicate whether or not the two constraints  $F_v < \bar{F}_v^{\text{min}}$  and  $F_v > \bar{F}_v^{\text{max}}$  are active for the video  $v$ . I then define:

$$\Gamma(\mathbf{x}, \mathbf{y}, D) = \sum_{v \in \mathcal{V}} x_v y_v \left( \frac{\alpha_v}{w_v D + \xi_v} + \beta_v \right) - \Lambda(\mathbf{x}, \mathbf{y}) \quad (3.5)$$

where

$$\Lambda(\mathbf{x}, \mathbf{y}) = \frac{\bar{R}_0}{\bar{H}} - \sum_{v \in \mathcal{V}} [(1 - x_v) \bar{F}_v^{\text{min}} - (1 - y_v) \bar{F}_v^{\text{max}}] \quad (3.6)$$

The pseudo-code of the video adaptation algorithm can be summarized as follows.

At each step, with given  $\mathbf{x}, \mathbf{y}$ , the solution of the equation  $\Gamma(\mathbf{x}, \mathbf{y}, D) = 0$  is derived. The resulting solution  $\tilde{D}$  is the fair distortion value to be assigned to the videos that have  $x_v = y_v = 1$ .

### 3.2.4 LTE Uplink Model

According to the LTE radio access model for the uplink, a time-slotted TDD SC-FDMA system is approached with  $S$  available sub-carriers divided in groups of 12 adjacent sub-carriers. The portion of frame composed by a single group of sub-carriers within one time slot (0.5ms) is defined as PRB, which is the elementary resource unit that can be allocated by the radio resource management unit within the eNodeBs.

---

**Algorithm 1** Pseudo code to solve problem (3.2).

---

```

1:  $y_v = 1, \forall v \in \mathcal{V}$ ;
2: repeat
3:    $x_v = 1, \forall v \in \mathcal{V}; \text{cond}_{HL} = \text{false}$ ;
4:   repeat
5:     Compute  $\tilde{D} : F(\mathbf{x}, \mathbf{y}, \tilde{D}) = 0$ ;
6:      $\text{cond}_{BL} = \text{false}$ ;
7:     for all  $v \in \mathcal{V} : x_v y_v = 1$  do
8:        $\tilde{F}_v = \frac{\alpha_v}{w_v \tilde{D} + \xi_v} + \beta_v$ ;
9:       if  $\tilde{F}_v < \tilde{F}_v^{\min}$  then
10:         $\tilde{F}_v = \tilde{F}_v^{\min}; x_v = 0; \text{cond}_{BL} = \text{true}$ ;
11:        break
12:      end if
13:    end for
14:    until  $\text{cond}_{BL}$ 
15:    for all  $v \in \mathcal{V} : x_v y_v = 1$  do
16:      if  $\tilde{F}_v > \tilde{F}_v^{\max}$  then
17:         $\tilde{F}_v = \tilde{F}_v^{\max}; y_v = 0; \text{cond}_{HL} = \text{true}$ ;
18:      end if
19:    end for
20: until  $\text{cond}_{HL}$ 

```

---

In this work, I focus on a single-cell scenario where users and eNodeBs are equipped with one and two antennas, respectively. Let  $h_{k,s}^{(i)}[n]$  be the channel gain between the  $k$ -th user and the  $i$ -th receiving antenna at the eNodeBs on sub-carrier  $s$  and time slot  $n$ . Each channel gain is modeled as a complex Gaussian random process (Rayleigh fading), in general correlated across sub-carriers and time slots, while no spatial correlation is assumed between the receiving antennas. Assuming maximum ratio combining at the receiver side, the normalized SNR of user  $k$  is defined on sub-carrier  $s$  and time slot  $n$ , as  $\gamma_{k,s}[n] = \frac{|h_{k,s}^{(1)}[n]|^2 + |h_{k,s}^{(2)}[n]|^2}{\sigma^2}$ , where  $\sigma^2$  is the noise power.

The radio resource allocation in LTE uplink systems operates in localized mode by allocating the available PRBs in a contiguous manner. By assuming an Minimum Mean Square Error (MMSE) receiver, the effective SNR experienced by user  $k$  over a set  $\mathcal{S}_k$  of sub-carriers forming  $\mathcal{G}_k$  adjacent PRBs in time slot  $n$  is given by:

$$\gamma_{k,\mathcal{S}_k}^{\text{eff}}[n] = \left( \frac{1}{\frac{1}{S_k} \sum_{s \in \mathcal{S}_k} \frac{p_{k,s} \gamma_{k,s}[n]}{1 + p_{k,s} \gamma_{k,s}[n]}} - 1 \right)^{-1} \quad (3.7)$$

where  $p_{k,s}$  is the power allocated to user  $k$  on sub-carrier  $s$ , and  $S_k$  is the cardinality of the set  $\mathcal{S}_k$ . Power control in LTE limits the amount of interference that User Equipment (UE) causes at base stations in neighboring cells. In this work, I consider the LTE uplink power control scheme described in [72].

By using a suitable Adaptive Modulation and Coding (AMC) scheme, the rate per unit bandwidth achievable by user  $k$  in time slot  $n$  when the set  $\mathcal{S}_k$  is allocated

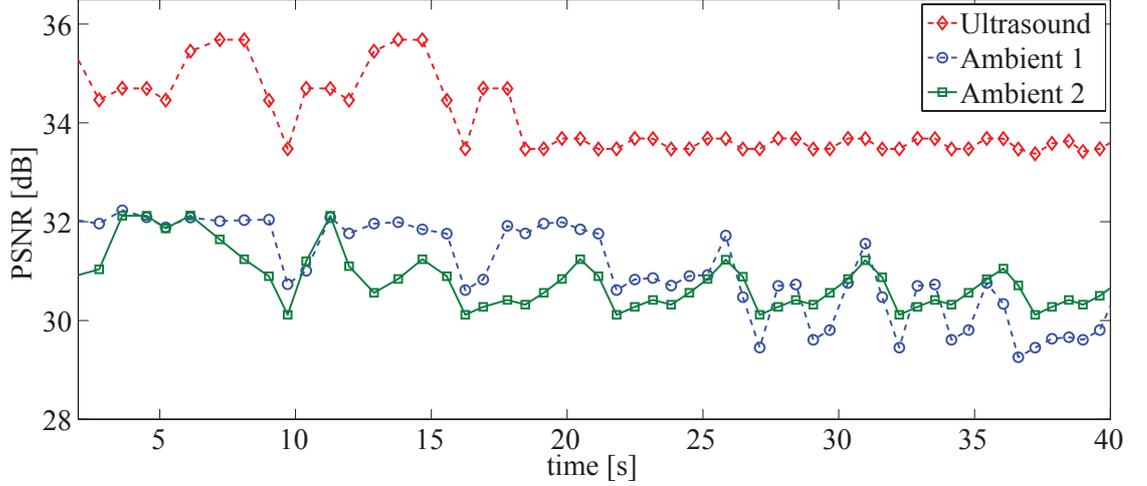


Figure 3.2: Average PSNR received at the hospital when a GBR  $\bar{R}_k = 4$  Mbps is guaranteed to the Ambulance.

can be evaluated with the following model:

$$r_k(\gamma_{k,S_k}^{\text{eff}}[n]) = S_k \cdot a_1 \log_2 \left( 1 + \frac{\gamma_{k,S_k}^{\text{eff}}[n]}{a_2} \right), \quad (3.8)$$

where  $a_1$  and  $a_2$  are two parameters called rate adjustment and SNR-gap respectively, depending on the specific AMC scheme adopted. Assuming turbo codes for forward error correction, I verified through numerical simulations that the modulation and coding schemes (MCSs) supported by LTE are characterized by a spectral efficiency well-approximated by the model (3.8). In particular, for this work the derived  $a_1 = 0.922$  and  $a_2 = 1.359$ , for a target Block Error Rate (BLER) =  $10^{-3}$  are taken into account. The MAC scheduler within the eNodeB allocates the rates to the users within the cell according to the solution of the following problem,

$$\max \sum_{k \in \mathcal{K}} R_k \quad (3.9a)$$

$$R_k \geq \bar{R}_k, \quad \forall k \in \mathcal{K}_1 \quad (3.9b)$$

$$R_k \geq \phi_k \left( \sum_{l \in \mathcal{K}} R_l - \sum_{l \in \mathcal{K}_1} \bar{R}_l \right), \quad \forall k \in \mathcal{K}_2 \quad (3.9c)$$

where  $R_k = \mathbb{E}_{\gamma}[r_k(\gamma)]$  is the ergodic rate of the  $k$ -th user and the parameters  $\phi_k \geq 0$ ,  $k \in \mathcal{K}_2$ ,  $\sum_{l \in \mathcal{K}_2} \phi_l = 1$ , define the required average share of throughput for

the best-effort users. The constraint (3.9b) accounts for the target rate requirement holding for the m-health user and the other GBR users within the cell. On the contrary, the rate allocated to non GBR users must satisfy the inequality (3.9c), which requires that the residual rate after serving all the GBR users (see the term in parenthesis at the right hand side) is assigned to the best effort users based on the  $\phi_k$ . The algorithms for deriving scheduling solution are obtained by following the procedure discussed in [73].

### 3.2.5 Numerical Results

An LTE-like access network is considered with frequency spacing  $\Delta B = 15$  kHz, transmission time interval equal to 1 ms and a maximum power budget of each user of 23 dBm. System bandwidth is set to 5 MHz, resulting in  $G=25$  allocable PRB. A total of  $K = 20$  users are uniformly distributed in a cell with resulting average SNR ranging from 5 to 28 dB. Specifically the m-health user experiences an average SNR of 13 dB and receives GBR traffic with values ranging from 2.5 to 7 Mbps. The other users are assumed as best-effort, i.e.  $\mathcal{K} = \mathcal{K}_2$ , with  $\phi_k = 1/K_2, \forall k \in \mathcal{K}_2$ . The radio channels for all users are modelled according to the ITU extended vehicular A model, with a doppler frequency of 70 Hz.  $N=2$  raw ambient videos are considered consisting of 300 frames with resolution 604x360 and one raw ultrasound sequence of 150 frames with resolution 640x480, each one acquired with a frame rate of 30 fps. Each sequence is looped in order to achieve 50 seconds long sequence and encoded with the Joint Scalable Video Model (JSVM) reference software [74] with one base layer and two enhancement layers. Each enhancement layer is split into five MGS layers with vector distribution [3 2 4 2 5]. GOP size and Instantaneous Decoder Refresh (IDR) period are set to 8 and to 32 frames, respectively. The resulting quality, in terms of average PSNR, ranges from 29 dB and 35 dB for the Ambient videos and from 32 to 40 for the Ultrasound video. The three parameters of model (3.1) are evaluated for each IDR period, resulting in adaptation intervals of about 1 sec. The video distortion weights  $w_v$  of the Ambient videos are set to 1, whereas Ultrasound weight is set to 2. Video playout deadline at the receiver is set to 200 ms. The overhead factor  $\bar{H} = 1$  is set. Fig. 3.2 shows the average PSNR at the receiver at each adaptation interval for the three video sequences when a GBR  $\bar{R}_0 = 4.5$  Mbps is provided to the m-health user. It is noticeable how the resulting qualities closely follow the selected video quality priorities providing a PSNR difference of approximately 3 dB between Ultrasound and Ambient videos. Fluctuation on the received PSNR are mainly due to the different spatial and temporal complexity of the scenes composing the video sequences. Such behaviour can also be appreciated in Fig. 3.3, where the average PSNR of each video sequences is plotted according to different GBR. As it can be noted by also considering Tab. 3.1, where the average rates of m-health and

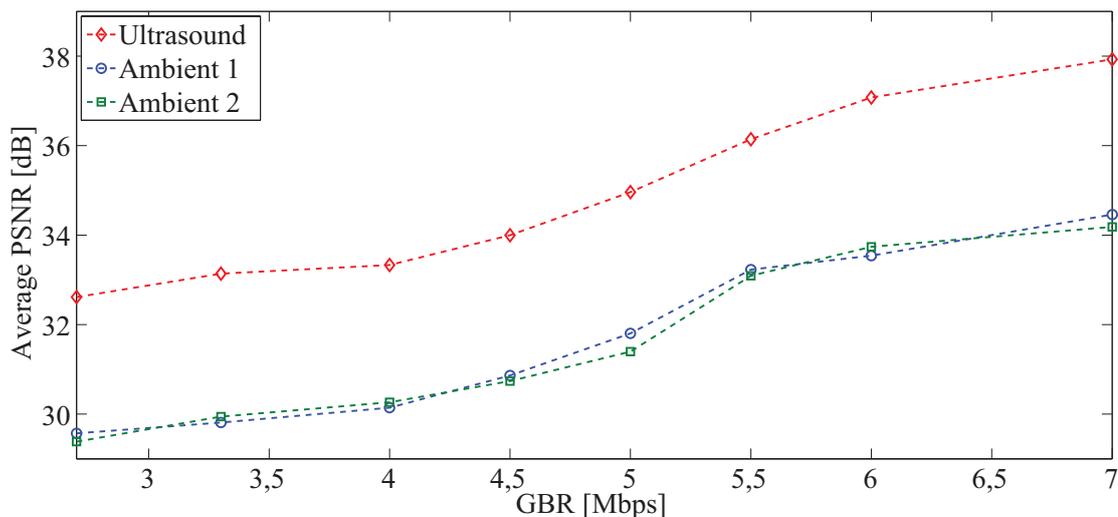


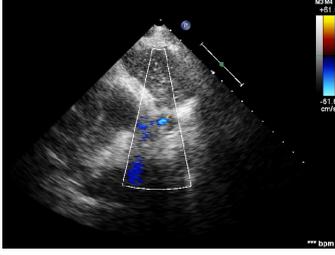
Figure 3.3: Average PSNR received at the hospital when different GBR are guaranteed to the Ambulance.

$\bar{R}_0$ [Mbps] (guaranteed)	2.5	3	4	4.5	5	6	7
$R_0$ [Mbps] (achieved)	2.7	3.3	4.2	4.5	5	6	6.8
$R_k, k \in \mathcal{K}_2$ [Mbps]	1.35	1.21	0.93	0.81	0.67	0.27	0.08

Table 3.1: M-health user GBR  $\bar{R}_0$ , average rate of e-health user  $R_0$  and average rate of the best-effort users  $R_k, k \in \mathcal{K}_2$ .

best-effort users are provided, our strategy decreases proportionally the quality of each video by reshaping the source rate, allowing reasonable throughput also for the best-effort users. On the other hand, an high GBR requirement, *i.e.*, 7 Mbps, allows to transmit almost entirely the highest enhancement layer, but almost all the physical resources are drained by the m-health user, thereby starving the best-effort users. Finally, in Fig. 3.2.5<sup>2</sup> a few examples of received frames are reported, as they result from the joint adaptation process. Frames (a)-(c) are extracted from to the ultrasound video sequence, while frames (d)-(f) from the ambient 1 video. Comparing the received frames with the original ones, it is noticeable that the proposed adaptation strategy is capable to achieve good visual results even when

<sup>2</sup>The authors would like to thank the "Green Cross" Public Assistance Association of Cesena (Italy) for their support in the acquisition of video sequences in a realistic on board ambulance scenario.



(a) Ultrasound original frame.



(b)  $\bar{R}_0 = 7$  Mbps. PSNR= 39.46 dB.



(c)  $\bar{R}_0 = 3.3$  Mbps. PSNR= 34.70 dB.



(d) Ambient original frame.



(e)  $\bar{R}_0 = 7$  Mbps. PSNR= 35.93 dB.



(f)  $\bar{R}_0 = 3.3$  Mbps. PSNR= 29.58 dB.

the overall available throughput for the m-health user is low ( $\bar{R}_0 = 3.3$  Mbps). At the same time, when the amount of available resources is higher ( $\bar{R}_0 = 7$  Mbps), the optimization strategy adaptively increases the final quality of all the videos. As a conclusion by observing that in all cases, the diagnostic ultrasound sequence is transmitted with higher quality with respect to ambient videos, according to the problem (3.2), enabling effective tele-diagnosis services.

### 3.3 Content/Context-Aware Multiple Camera Selection and Video Adaptation For The Support of mHealth Services

In this Section I focus on the problem of delivering multiple health-related real-time video streams from an emergency scenario to a remote hospital by exploiting the uplink of an LTE wireless access network, in order to support efficient mHealth tele-consultation services. As described in Sec. 1.2, the transmission of health-related information is a challenging task, due to the variability and the limitations of the mobile radio link, the different qualities of the visual representations of the cameras and the heterogeneous end-to-end quality requirements of the contents to be delivered. The following solution is based on: (i) the application of the

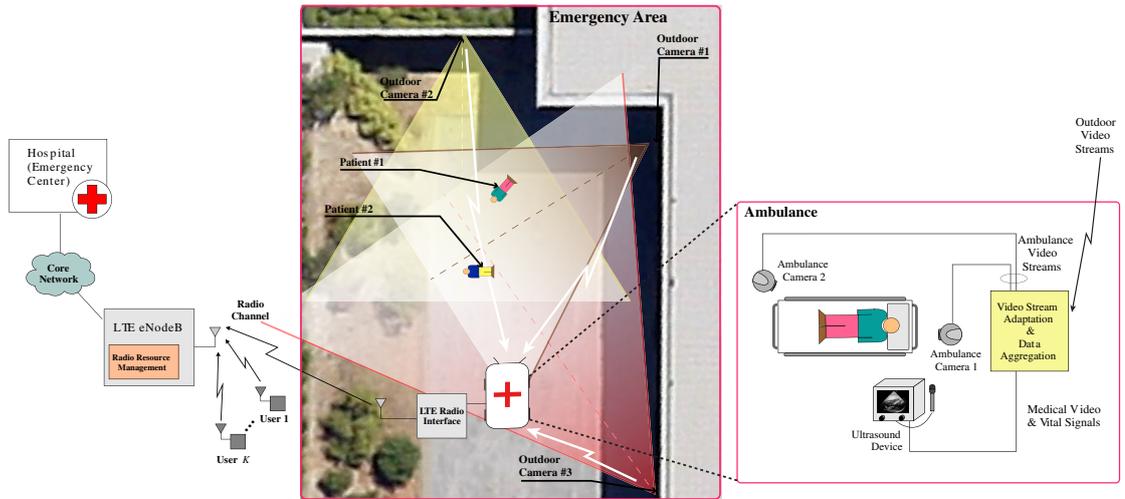


Figure 3.4: The proposed mHealth architecture for emergency scenarios.

context-aware camera selection algorithm (CRSS), presented in Sec. 2.2.2, which selects among the cameras deployed in the emergency scenario one or more video sources taking into account specific ranking criteria mainly related to the quality of the visual representation of the object of interest; (ii) a content-aware technique for the transmission of multiple scalable videos that jointly considers video aggregation and adaptation at the application layer of the transmitting equipment and takes into account the different quality requirements of diagnostic and ambient videos. Numerical results show that the proposed strategy permits to achieve a good end-to-end quality for both the diagnostic and the ambient videos even in the presence of rate limitations and fluctuations in the wireless link, due to the channel variations and the traffic load inside the LTE cell. When the wireless link capacity decreases, the proposed strategy appropriately discards the videos coming from the cameras providing the lowest visual quality, according to the camera ranking results, and, at the same time, adapts the rate of the transmitted videos to provide the requested quality with priority to diagnostic content [1].

### 3.3.1 System Architecture

Real-time video streams are acquired by a camera network deployed in the emergency area with monitoring functionalities and one diagnostic sequence is acquired directly on board of the ambulance. The multimedia flows are processed in real-time and multiplexed with the other medical information by the video adaptation solution. The mHealth scenario addressed in this paper is depicted in Fig. 3.4. In the considered emergency area one or multiple injured persons need immedi-

ate medical assistance. From the LTE network perspective, the ambulance (also called m-health user) competes for radio resources with other  $K$  users within the cell, indexed by the set  $\mathcal{K}$ , subdivided into  $K_1$  GBR users and  $K_2$  best-effort users, indexed by the sets  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , respectively. The eNodeB tries to guarantee the transmission rates  $\bar{R}_0$  to the m-health user.

In this scenario, two different phases of the first-aid operations are considered. In the first stage, the ambulance collects and sends up to three outdoor ambient videos in which paramedics are providing the first aid to the patient. In the second stage a patient, who has been loaded on the ambulance, is being visited with a ultrasonography. In this phase, the ambulance collects and sends up to two ambient indoor videos and one diagnostic video sequence. Each video has been acquired at the hospital of Perugia, Italy, in realistic outdoor and on-board ambulance scenarios. In each stage, the multiple aggregated videos are sent through the available LTE radio access network to the emergency management center at the hospital, where specialized medical staff can follow the first-aid operations, coordinate the intervention and acquire the health-state information necessary to prearrange the treatment at the hospital.

The video adaptation unit on board the ambulance manages the inherently different priorities of the data flows generated by the m-health user. In particular, it optimally adapts the SVC-encoded streams, in order to deliver the ultrasonography information with sufficiently high quality and the set of ambient videos tuned according to quality fairness criteria. To this purpose, it generates an aggregated throughput adapted to the radio channel and cell traffic conditions, according to the amount of resources assigned by the eNodeB to the m-health user.

### 3.3.2 Multi Camera Ranking

In this Section, the novel camera selection algorithm presented in Sec. 2.2.2 is applied, which selects one or more videos taking into account a specific ranking criteria, mainly related to the quality of the visual representation of the object of interest. According to Fig. 3.4, a certain area has to be monitored with a multi camera system connected to the ambulance through local wireless links. The goal of the ranking consists in determining which camera provides the image of the object with the best resolution. More precisely, the proposed camera ranking procedure is based on a quality metric capable to combine the size of the object and the pixel density on the image plane. The objects size projected on the image plane depends on the distance of the objects to the camera: the closer are the objects, the bigger are their projections on the image plane. Moreover, the pixel density depends on the resolution of the single camera: the higher is the resolution, the more accurate is the resulting image. The best image quality is obtained by the camera that guarantees the best trade-off in terms of distance and

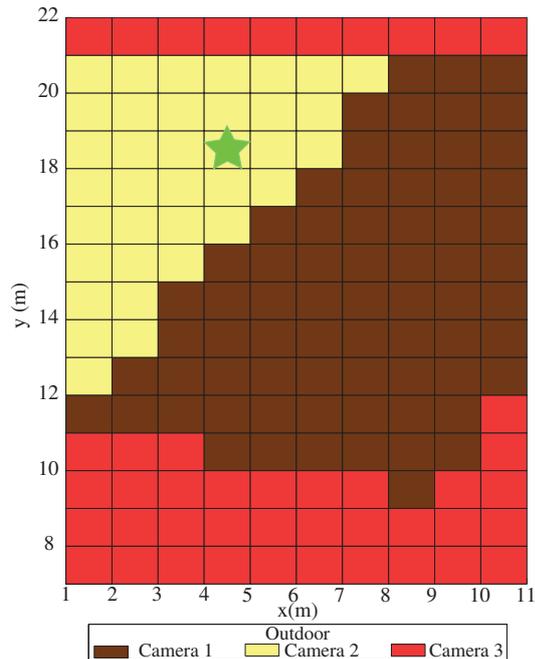


Figure 3.5: The camera ranking algorithm results for the whole emergency area.

resolution [1]. It is assumed that all the involved video devices are equipped with GPS or Real-Time Location System (RTLS) in order to automatically identify the camera pose and to self-determine the orientation of each camera itself. It is remarkable that the proposed camera ranking algorithm is suitable for static or low-motion scenes; further camera selection strategies that take into account the acquisition frame rate and the average object velocity in the scene are proposed in Sec. 2.2.3. Considering the scenario in Fig. 3.4 the emergency area is monitored with 3 cameras (*Outdoor camera 1*, *Outdoor camera 2*, *Outdoor camera 3*) placed in different positions. Moreover, the monitored area is discretized in small sub-area units; for each of them the proposed algorithm selects the camera that describes the given sub-area with the highest number of pixel per area. For the sake of clarity the camera selection results is graphically showed for a set of sub-areas lying on the same plane orthogonal to the z-axis. In addition, all the involved cameras have the same resolution. The corresponding camera selection output is reported in Fig. 3.5, where different sub-area colors correspond to different cameras. In particular, the sub-volume highlighted in Fig. 3.5 (in correspondence to the position of one the patients), which refers to the *Outdoor camera 2*, is capable to provide the object description with the highest resolution in terms of pixel per object unit area while *Outdoor camera 1* and *Outdoor camera 3* are ranked at the second and third position respectively.

### 3.3.3 Video Coding And Adaptation

The video sequences are organized in sets of consecutive frames called GOPs, which are encoded in scalable streams. From a scalable stream it is possible to extract a sub-stream in order to meet a given frame rate, resolution and quality, according to three scalability methods supported by SVC. I focus on quality, *i.e.*, SNR, scalability, which is obtained by enhancing the quality of the video stream with the addition of refinement layers. Two different possibilities are now available in the SVC standard [54], namely the CGS and the MGS. In this work, I focus on MGS scalability, which provides finer granularity with respect to CGS coding.

Rate adaptation techniques dynamically adapt the amount of transmitted information to the available channel bandwidth by taking into account the content of the videos and its impact on the end-user quality. R-D models enable to predict the minimum bit rate (in bit/s, or bps) required to achieve a target distortion, which can be evaluated according to a particular objective video quality metric, *e.g.*, MSE, PSNR or SSIM metrics. The SSIM video quality metric is considered, which has been recently shown to provide a higher correlation with subjective tests when assessing the diagnostic quality of ultrasonography video sequence [75].

Let us consider the  $v$ -th SNR-scalable video ( $v \in \mathcal{V}$ , where  $\mathcal{V}$  is the set of the  $N$  videos considered in each operation step), resulting from the encoding of the set of pictures  $\mathcal{I}_v$ . The minimum information rate in bps,  $F_k(d_{e,v})$ , required to transmit the  $e$ -th sub-stream with the given expected distortion  $d_{e,v}$ , is generally a function of discrete values. Following the approach in [71] and recently extended in [65], the expected R-D relationship can be modeled by using a parametric function  $F_v(D)$  of a continuous variable  $D$ , defined over a limited interval, *i.e.*,  $D \in [D_v^{\text{hl}}, D_v^{\text{bl}}]$ . Within SVC,  $D_v^{\text{bl}}$  and  $D_v^{\text{hl}}$  are the expected distortions of the set of pictures  $\mathcal{I}_v$  after decoding the base layer (with rate  $F_v^{\text{min}}$ ) and the highest enhancement layer (with rate  $F_v^{\text{max}}$ ), respectively. Although this framework holds for any monotonic strictly decreasing function  $F_v(D)$ , the following parametric model is considered:

$$F_v(D) = \alpha_v e^{\xi_v D} + \beta_v, \quad D \in [1 - Q_v^{\text{hl}}, 1 - Q_v^{\text{bl}}], \quad \xi_v < 0, \quad (3.10)$$

where  $D = 1 - Q$ ,  $Q$  is the average SSIM computed between the original and the reconstructed set of pictures, and  $Q_v^{\text{bl}}$ ,  $Q_v^{\text{hl}}$  are the average SSIM of the set of pictures  $\mathcal{I}_v$  after decoding the base layer and the highest enhancement layer, respectively. The three parameters  $\alpha_v$ ,  $\xi_v$  and  $\beta_v$  depend on the temporal and spatial characteristics of the set of pictures  $\mathcal{I}_v$  and on the frame rate. A simulation campaign has been carried out to validate the model (3.10). The results, which will be reported on an extended version of this paper, have shown almost perfect correlation with a Pearson coefficient always higher than 0.9999 for each set of encoded pictures.

Let us assume that a bearer with a GBR  $R_0$  has been granted to the ambulance, after a suitable negotiation with the LTE network, in order to support the emergency m-health services. Such value  $R_0$  is eventually updated every  $W$  seconds in case of critical cell-load conditions or bad channel conditions for the ambulance. The m-health user exploits this guaranteed bandwidth to deliver the best video quality according to priority and fairness constraints defined for the different videos. This is obtained through a dynamic rate adaptation strategy, consisting in maximizing the overall video quality while minimizing a weighted quality difference among the different videos under minimum and maximum rate constraints. This strategy has been first proposed by the authors in [65] by addressing the problem for both empirical and semi-analytical R-D model and then extended in Sec. 3.2 to consider the different priorities of the video sequences, while using the MSE as video distortion metric [3]. In [65] it has been shown that, when the parametric R-D model is sufficiently accurate, it can be used to relax the multi-objective optimization problem leading to a much simpler constraint satisfaction problem. Here, this approach is followed by defining an adaptation strategy aimed at deriving the transmission rates, as the solution of the following set of equations and constraints:

$$\sum_{v \in \mathcal{V}} HF_v = R_0, \quad (3.11a)$$

$$\Delta(D_i, D_j; w_i, w_j) = 0 \quad \forall i, j \in \mathcal{V}, i > j \quad (3.11b)$$

$$F_v^{\min} < F_v < F_v^{\max} \quad \forall v \in \mathcal{V} \quad (3.11c)$$

where  $\mathcal{V}$  is the set of videos handled by the ambulance for e-health emergency services in each operation step,  $H$  is the estimated overhead introduced at the different layers of the network architecture,  $w_v$ , with  $v \in \mathcal{V}$ , are the weights used to account for the different priorities, and  $\Delta(D_i, D_j; w_i, w_j)$  is the distortion-fairness metric for each pair of videos, defined as [65]:

$$\Delta(D_i, D_j; w_i, w_j) = \begin{cases} 0 & \text{if } (D_i = D_i^{\text{hl}} \wedge w_j D_j < w_i D_i) \vee \\ & (D_j = D_j^{\text{hl}} \wedge w_j D_i < w_i D_j) \\ 0 & \text{if } (D_i = D_i^{\text{bl}} \wedge w_j D_j > w_i D_i) \vee \\ & (D_j = D_j^{\text{bl}} \wedge w_j D_i > w_i D_j) \\ |w_i D_i - w_j D_j| & \text{otherwise.} \end{cases} \quad (3.12)$$

where  $\wedge$  and  $\vee$  are "AND" and "OR" operators, respectively. As an example, by setting the weights equal to 2 for the most important videos, *e.g.*, ultrasound sequence, and equal to one for the other less important videos, *e.g.*, ambient sequences, the adaptation module at reasonable GBR values will be able to provide

an SSIM value higher than 0.95 to the ultrasound video sequence, as long as the other ambient videos are supported with an SSIM not lower than 0.9. In this case, the perceived quality of the medical video is suitable to diagnostic purposes [75], while ambient videos achieve a reasonable but lower quality. It is remarked that, as the GBR decreases, also the difference in quality between ambient and diagnostic videos will decrease and *vice versa*.

The optimal algorithm proposed in [3] to solve problem (3.11) according to a MSE-based R-D model can be suitable extended to consider the SSIM-dependent R-D function considered here, by replacing the functions in (5) and (6) in Sec. 3.2 as follows:

$$\Gamma(\mathbf{x}, \mathbf{y}, D) = \sum_{v \in \mathcal{V}} \left[ x_v y_v \left( \alpha_v e^{w_v \xi_v D} + \beta_v \right) + (1-x_v) \bar{F}_v^{\min} + (1-y_v) \bar{F}_v^{\max} \right] - \frac{R_0}{H}. \quad (3.13)$$

I refer to [3] for further details.

### 3.3.4 Numerical Results

In the following simulations, LTE-like access network is considered with sub-carrier spacing  $\Delta B = 15$  kHz and system bandwidth set to 5 MHz. A total of  $K = 15$  users with a maximum per-user power budget of 23 dBm are uniformly distributed in a cell of 300 m, resulting in an average SNR ranging from 5 to 28 dB. More specifically the ambulance user experiences an average SNR of 13 dB and receives from the LTE access network a GBR  $R_0$  with values negotiated in the range from 1 to 5.5 Mbps. The other users are assumed best-effort, *i.e.*,  $K = K_2$ . The radio channel for all users is modeled according to the International Telecommunications Union (ITU) extended pedestrian A model, with a Doppler frequency of 5 Hz. In the first stage the ambulance send up to three outdoor ambient videos, denoted as *Outdoor Camera 1*, *Outdoor Camera 2*, *Outdoor Camera 3* with resolution 720p (1024x720). In the second phase, the ambulance collects and sends two ambient indoor videos and one diagnostic video sequence, namely *Ambulance Camera 1*, *Ambulance Camera 2* and *Ultrasound*, with VGA resolution (640x480). All videos are acquired with a frame rate 30 fps.

Each video sequence is encoded with the JSVM reference software [74] with one base layer and three SNR enhancement layers. MGS is considered to encode the quality layers and both enhancement frames and base layer key-picture frames are used for motion compensation, thereby limiting the drift issue during the adaptation process. The Quantization Parameter (QP) of the ambient videos base layer (indoor and outdoor) is set to 46, while the ultrasonography video base layer is encoded with QP equal to 38. The QP difference between enhancement layers for all videos is set to 4. The GOP size and the IDR period are here set to

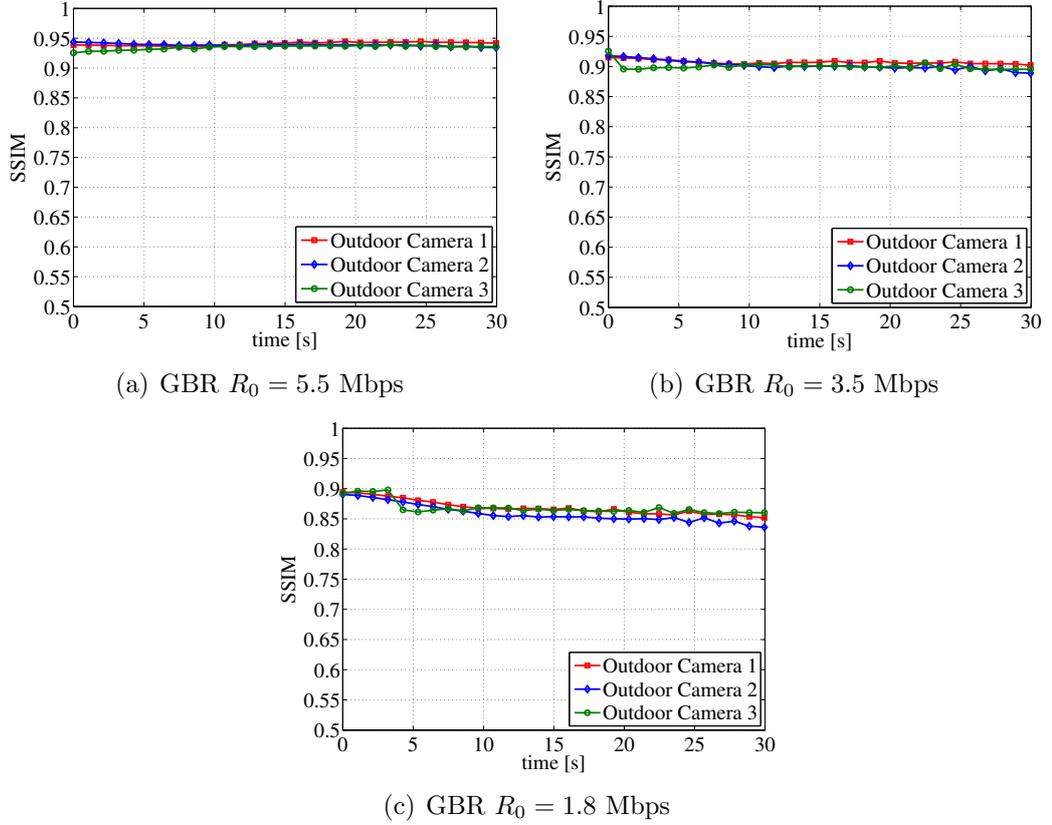


Figure 3.6: Received SSIM averaged over each adaptation interval, when transmitting the three outdoor video sequences in the first stage, for different values of GBR granted to the ambulance. The quality weights are set to 1 for all videos.

8 and to 32 frames, respectively. After encoding, the resulting quality in terms of the average SSIM, ranges approximately from 0.82 to 0.95 for the ambient videos and from 0.89 to 0.98 for the ultrasound video. The three parameters of model (3.10) are evaluated for each IDR period, resulting in adaptation interval of about 1 sec. The video distortion weights  $w_v$  of the ambient videos are set to 1, whereas the weight of the ultrasound video is set to 2. Finally, video play-out deadline at the receiver is set to 200 ms and the overhead factor is set to  $H = 1$ .

Fig. 3.6 shows the SSIM at the receiver averaged over each adaptation interval when transmitting the three outdoor video sequences in the first stage, with different GBR values provided to the ambulance, *i.e.*,  $R_0 = 5.5$  Mbps (a),  $R_0 = 3.5$  Mbps (b) and  $R_0 = 1.8$  Mbps (c). It is noticeable how the proposed strategy provides approximately the same SSIM to each ambient video at each adaptation interval. As the GBR value granted to the ambulance decreases, the rate of each video is reshaped accordingly, thereby providing a graceful degradation while

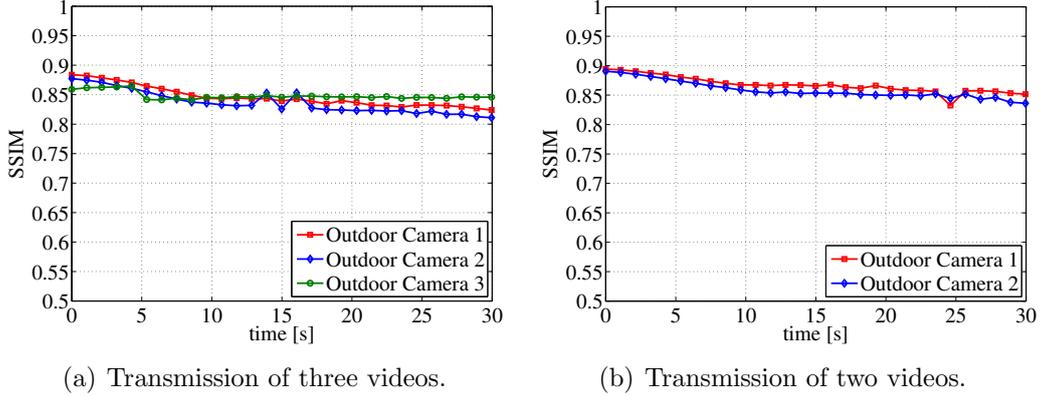


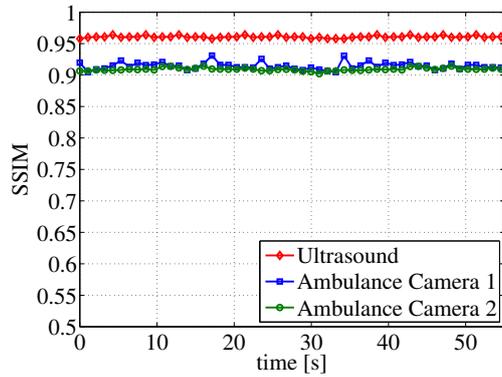
Figure 3.7: SSIM averaged over each adaptation interval received at the hospital, when transmitting all the three (a) and two (b) outdoor video sequences in the first stage. GBR  $R_0 = 1.0$  Mbps. The quality weights are set to 1 for all videos.

maintaining quality fairness.

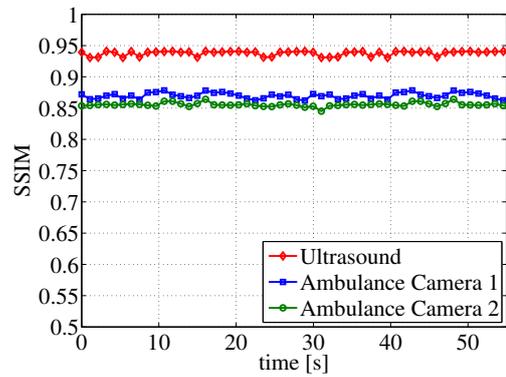
When the GBR granted is low, as in the case of  $R_0 = 1.8$  Mbps (c), the SSIM decreases due to the increasing spatial and temporal complexity of the specific scenes composing the video sequences. Fluctuations on the SSIM are due to the larger gap between continuous and discrete solution, when high resolution video sequences are considered.

When the GBR granted to the ambulance is set to  $R_0 = 1$  Mbps, the transmission of all three videos with the same weights leads to an SSIM close to 0.8 as shown in Fig. 3.7(a), which results to an unacceptable received quality. In this case, the ambulance exploits the video ranking proposed in Sec. 3.3.2 to increase the quality weights of the most important videos or it may opt for a hard decision by transmitting only the video with the highest ranking value, dropping low priority sequences until a reasonable quality is received. An example of the second case is shown in Fig. 3.7 (b), where the lowest ranked video (*Outdoor Camera 3*), is no longer transmitted. As a consequence, the SSIM received at the hospital for the highest ranked video sequence increases, leading to an SSIM not lower than 0.85.

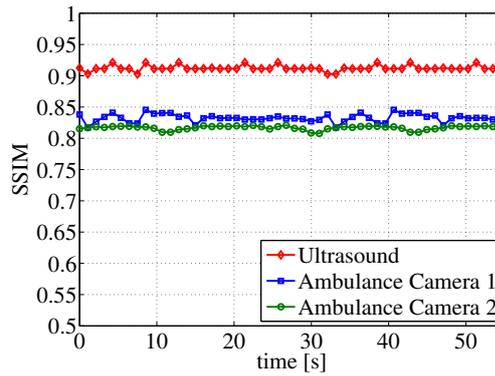
Fig. 3.8 shows the SSIM at the receiver averaged over each adaptation interval in the second stage, when transmitting the two indoor ambient video sequences together with the diagnostic ultrasonography video. The same GBR values provided to the ambulance in the first stage are considered. It is also remarked how the resulting qualities closely follow the selected video quality priorities providing an SSIM difference that decreases proportionally as the individual SSIM increases, thereby delivering the diagnostic ultrasound sequence with higher quality with respect to ambient videos and enabling effective tele-diagnosis services.



(a) GBR  $R_0 = 5.5$  Mbps



(b) GBR  $R_0 = 3.5$  Mbps



(c) GBR  $R_0 = 1.8$  Mbps

Figure 3.8: SSIM averaged over each adaptation interval received at the hospital, when transmitting the two indoor ambient video sequences and the ultrasonography video in the second stage, for different values of GBR granted to the ambulance. The quality weights are set to 1 for the ambient videos, and 2 for the ultrasonography video.

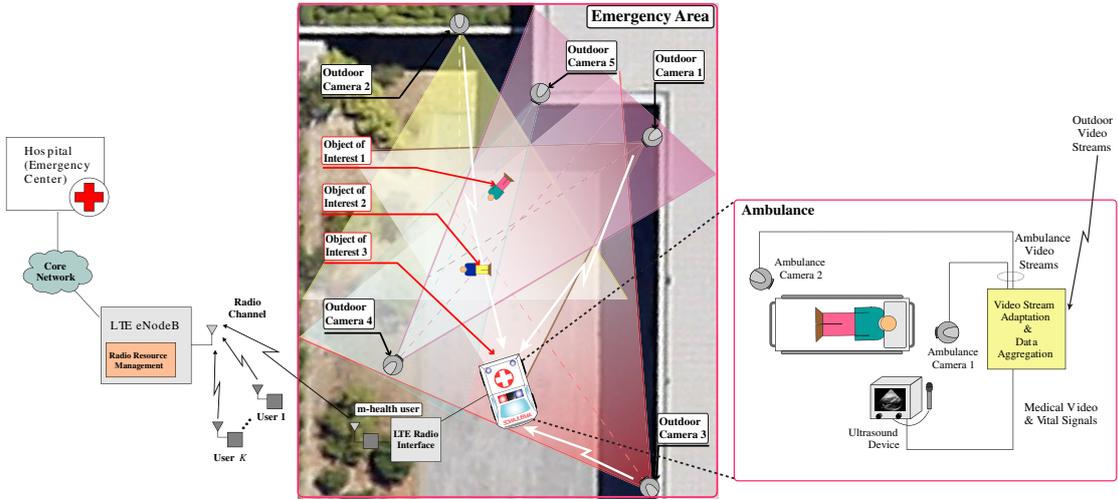


Figure 3.9: The proposed m-health architecture for emergency scenarios.

### 3.4 An Optimization Framework for Content/Context-aware SVC Multiple Video Delivery In mHealth Emergency Applications

Transmission of health-related information from ambulances to remote hospitals allows specialists to perform preliminary diagnostic analysis and to prepare the hospitalization adequately, as well as to provide specialized support to local staff. The presence of a 4G radio access network can be exploited to establish a communication link with the emergency area, according to the recent models for next-generation public safety networks [76]. The selected videos (ambient and/or diagnostic) are collected and aggregated to coordinate the transmissions to and from the hospital [77]. The possibility to adapt the video encoding to the current transmission conditions becomes particularly important in the context of emergency m-health, because mobile links are usually characterized by limited and variable bandwidth. Among video coding standards, SVC [54] conjugates good compression efficiency with high flexibility in rate adaptation. For this reason several solutions have been recently proposed for e-health applications based on SVC [78]. This work is the extended version of the approach proposed in Sec. 3.3: a camera ranking technique is firstly proposed, which permits to select one or more cameras taking into account specific quality criteria mainly related to the visual representation of the object of interest. The proposed context-aware approach has been developed with the objectives (i) to maximize the QoE for the final user, (ii) to efficiently manage distributed hardware resources, and (iii) to optimize the usage

of radio resources for video transmissions. Then, a novel solution is described for the transmission of the selected videos from the emergency area, based on joint video adaptation and aggregation directly performed at the application layer of the processing equipment within the ambulance. In our solution the adaptation is designed to optimize quality and fairness by exploiting the information on the available rate assigned by the LTE e-nodeB. It is shown that the proposed strategy permits to achieve a good end-to-end quality even in the presence of rate limitations and fluctuations due to the wireless channel and intense traffic within the LTE cell. When the channel conditions deteriorate, the adaptation strategy appropriately discard the cameras which provide the lowest visual quality based on the camera ranking results[4].

### 3.4.1 System Architecture

The m-health scenario addressed in this paper is depicted in Fig. 3.9. In the considered emergency area one or multiple injured persons need immediate medical assistance. An ambulance, equipped with multiple cameras and diagnostic devices, is present on the scene of the accident. Multiple real-time video streams are acquired both on-board and outside the ambulance, which is responsible for locally processing the collected information and transmitting the most significant part of them to the remote hospital. External videos are captured by a camera network deployed for monitoring purpose, including stationary devices and, possibly, mobile cameras worn by the medical staff. In this work, it is assumed that the position and attitude of each camera are known in real-time, *e.g.*, by means of inertial units mounted on the devices [79], GPS receivers or other RTLS [80, 81, 82]. Modern equipment for emergency teams, in fact, are always more frequently capable to provide location information about people and objects on the field [83], in order to improve the efficiency and the safety of the operations [84, 85, 86]. Multimedia flows are collected by the ambulance, processed in real-time and multiplexed with other medical information. Then, a joint stream adaptation is performed on the ambulance taking into account the camera ranking results, the radio channel and cell traffic conditions, as well as the amount of resources assigned by the LTE/Evolved Packet Core (EPC) network. From the LTE network perspective, the ambulance (in the following also called m-health user) competes for radio resources with other  $K$  users within the cell, indexed by the set  $\mathcal{K}$ , subdivided into  $K_1$  GBR users and  $K_2$  best-effort users, indexed by the sets  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , respectively. The e-NodeB tries to guarantee the transmission rates  $\bar{R}_0$  to the m-health user<sup>3</sup> and  $\bar{R}_k$  to the  $k$ -th GBR user, with  $k \in \mathcal{K}_1$ , while the throughput to best-

---

<sup>3</sup>Note that here, as well as in the rest of the paper, the m-health GBR user is indicated with the subscript 0.

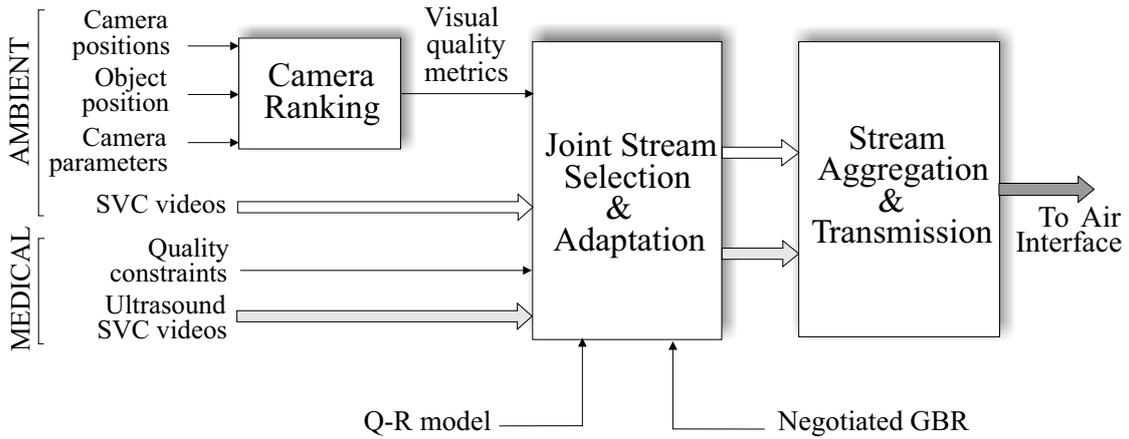


Figure 3.10: Block diagram reporting the exchange of information in the proposed multi-camera management system.

effort users is provided fairly, according to the remaining capacity available after allocating all GBR users [87, 3]. It is assumed that the first-aid operations take place in two different phases. In the first phase, the ambulance collects and sends up to  $N$  outdoor ambient videos in which paramedics approach the patients to provide first assistance. In the second stage a patient is loaded on the ambulance, where an ultrasound examination is performed to check his/her health status. In this phase, the ambulance collects and sends up to two ambient indoor videos and one diagnostic video sequence. The video adaptation unit on-board the ambulance manages the inherently different priorities of the data flows generated by camera network and diagnostic equipment. In particular, it optimally adapts the SVC-encoded streams, in order to deliver the ultrasound video (when present) with sufficiently high quality and a set of ambient videos tuned according to ranking and fairness criteria. In Fig. 3.10 the algorithmic architecture of the proposed solution is summarized, highlighting the exchange of information among the different logical units. As it can be noted, the Camera Ranking Algorithm (CRA) process the side information relevant to positions and parameters of the ambient cameras, providing as output a set of visual quality metrics. Based on these values, the following unit selects the optimal set of ambient video to transmit, while guaranteeing a sufficient video quality for high priority diagnostic streams (when present). As it will be discussed in the following sections, video selection is jointly realized with real-time stream adaptation.

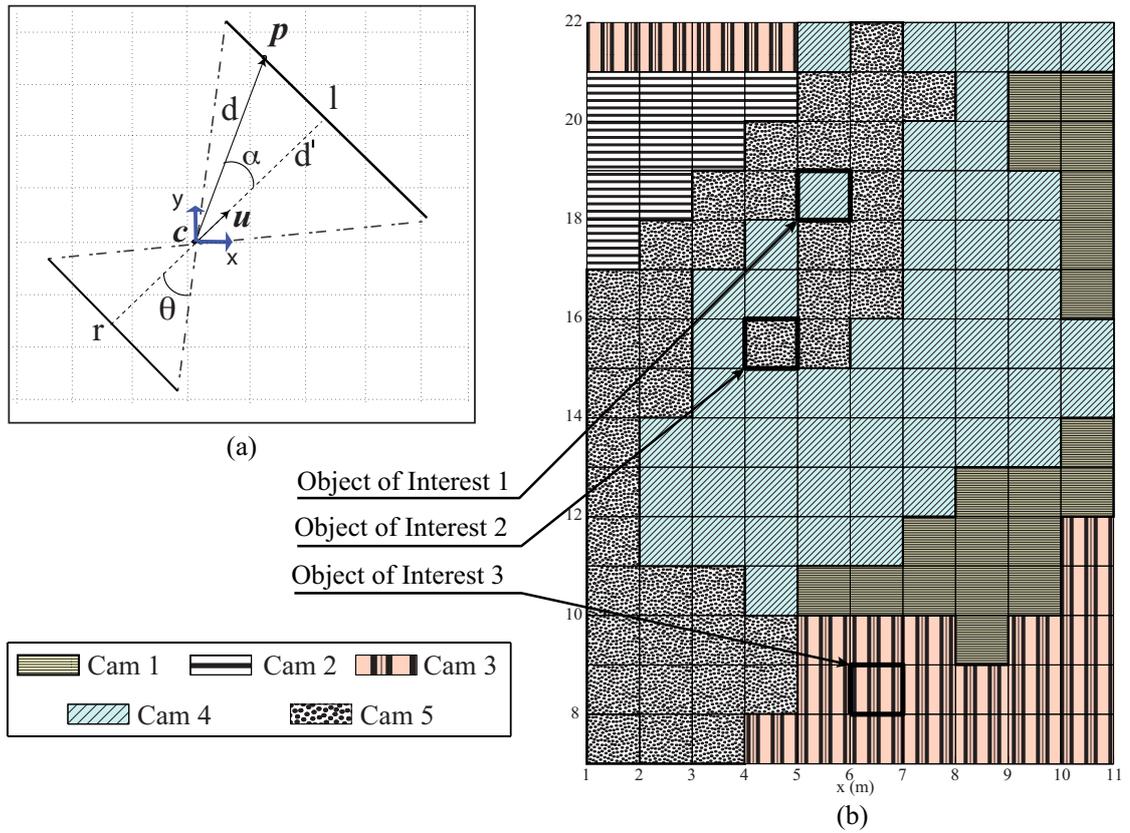


Figure 3.11: (a) the 2D pinhole camera model. In evidence the FoV, the camera orientation and the position of an object of interest. (b) the camera ranking algorithm results for the whole emergency area.

### 3.4.2 Multi Camera Ranking

In this Section, the novel camera selection algorithm presented in Sec. 2.2.2 is applied, which selects one or more videos taking into account a specific ranking criteria, mainly related to the quality of the visual representation of the object of interest. According to Fig. 3.9, a certain area has to be monitored with a multi camera system connected to the ambulance through local wireless links. The proposed ranking technique is based on geometrical considerations, as it is assumed that the visual quality mainly depends on the orientation of each camera and the 3D viewing volume that can be captured. First of all, through the preliminary calibration process, the intrinsic parameters of every camera are determined. The knowledge of the focal length, in combination with the aspect ratio of the image plane, permits to calculate the different FoVs. Based on camera FoV and spatial orientation, it is possible to know which portion of the area of interest is monitored

Notation	Description	Notation	Description
$\mathcal{K}, K$	Set and total number of users	$N$	Total number of cameras
$\bar{R}_0$	Target GBR value	$F_v$	Source Rate of video $v$
$q_v$	Visual quality of video $v$	$p_v$	Priority of video $v$
$w_v$	Weight of video $v$	$a_v, b_v, c_v$	Utility parameters for video $v$
$U_v^{\text{target}}$	Target utility of video $v$	$U_v$	Utility of video $v$
$F_v^{\text{bl}}$	Base layer rate for video $v$	$F_v^{\text{hl}}$	Maximum source rate for video $v$
$\alpha$	Shaping value	$H$	Overhead factor
Acronym	Full name	Acronym	Full name
CRA	Camera Ranking Algorithm	BS	Base Station
SVC	Scalable Video Coding	Q-R	Quality-rate
SSIM	Structural Similarity	GOP	Groups of Pictures
GBR	Guaranteed Bit-rate	SNR	Signal-to-noise Ratio
IDR	Instantaneous Decoder Refresh		

Table 3.2: List of most used symbols and acronyms.

by the available devices. The ranking procedure consists in ordering the cameras according to a visual quality metric that combines the position of the object of interest with respect to the camera and the pixel density on the image plane. The size of the object projection on the image plane depends on the distance between the object and the camera: the closer is the object, the wider is its projection. At the same time, the image quality, in terms of perceivable details, increases with the camera resolution. Hence, the highest visual quality for the object of interest is provided by the camera that guarantees the best combination of distance and resolution, here measured as the amount of pixels required to represent a unit area surface orthogonal to the camera orientation placed at the considered distance. Now, the complete scenario depicted in Fig. 3.9 is considered. The emergency area (in this case monitored with 3 cameras) is partitioned in sub-areas, as depicted in Fig. 3.11b. For each sub-area, the proposed algorithm ranks the available cameras based on the number of pixels used to represent a unit area surface placed at the corresponding distance. In Fig. 3.11b, the top ranking camera it is showed for each sub-area on a plane perpendicular to the  $z$ -axis, discriminating the devices by assigning different filling patterns. The object of interest 1 (in our simulation corresponding to an injured person on the ground) is optimally captured by the outdoor *Camera 4*, while the best ranking acquisition devices for the objects of interest 2 and 3 are *Camera 5* and *Camera 3*, respectively. In the following, the ranking value  $q_v$  denotes the number of pixel per unit area relevant to the  $v$ -th camera. To simplify the notation, it can be assumed *w.l.o.g* that, for the considered object of interest, the cameras are already numbered according to the ranking values, that is  $q_i > q_j, \forall i > j$ . Finally, defining  $\mathcal{V}_x$  as the set of the  $x$  highest ranking cameras, we have  $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \dots \subset \mathcal{V}_N$ .

### 3.4.3 Joint Video Selection And SVC Adaptation

#### Quality-Rate Model For SVC

The SVC standard is attractive for application involving multi-camera emergency scenarios, as it permits real-time and distributed stream adaptation, based on the encoding of video sequences into scalable streams. From a scalable stream it is possible to extract a sub-stream in order to meet a given frame rate, resolution and video quality, according to the different scalability methods supported by SVC. I focus on video quality scalability, also known as SNR scalability. SNR scalability is obtained by enhancing the quality of the video stream with the addition of refinement layers. Two different possibilities are available in the SVC standard [54], namely the CGS and the MGS. MGS provides finer granularity with respect to CGS coding and it is obtained by dividing a CGS layer into up to 16 MGS layers. In this work, I focus on MGS scalability. Rate adaptation techniques dynamically adapt the amount of transmitted information to the available channel bandwidth by taking into account the context and the content of the video cameras and its impact on the end-user video quality. Quality-rate (Q-R) models enable to predict the minimum bit rate (in bit/s or bps) required to achieve a target video quality, which can be evaluated according to a particular objective video quality metric, *e.g.*, MSE, PSNR or SSIM metrics [88]. Here, the SSIM is considered as video quality metric, which has been recently shown to provide a higher correlation with subjective tests when assessing the diagnostic quality of ultrasonography video sequence [75]. The SSIM index has values ranging between -1 and 1, increasing with the video quality.

Each video sequence is organized in sets of consecutive frames named Groups of Picturess (GOPs). The interval between two consequent I-frames, also known as IDR period, is here assumed as a multiple of the GOP size. Let us consider the SNR-scalable video resulting from the encoding of the set of pictures  $\mathcal{I}_v$ , captured by the camera  $v \in \mathcal{V}_N$ . The quality of this video, indicated by the utility  $U_v$ , depends on the coding rate  $F$ , *i.e.*,  $U_v = U_v(F)$ , and generally is a discrete valued function, as the rates corresponding to available SVC layers belong to a discrete set. Following the approach in [71] and recently extended in [65][89], the expected Q-R relationship can be modeled by using a parametric function  $U_v(F)$  of the continuous variable  $F$ , defined over the limited interval  $F \in [F_v^{bl}, F_v^{hl}]$ , where  $F_v^{hl}$  and  $F_v^{bl}$  represent the two limit rates corresponding to all the available enhancement layers and the sole base layer, respectively. Clearly, encoding the set of pictures  $\mathcal{I}_v$  with rate  $F_v^{hl}$  provides the maximum utility  $U_v^{hl}$ , while  $U_v^{bl}$  is the utility of the base layer. Although the proposed framework holds for any monotonic strictly increasing and invertible function  $U_v(F)$ , in this work the following parametric

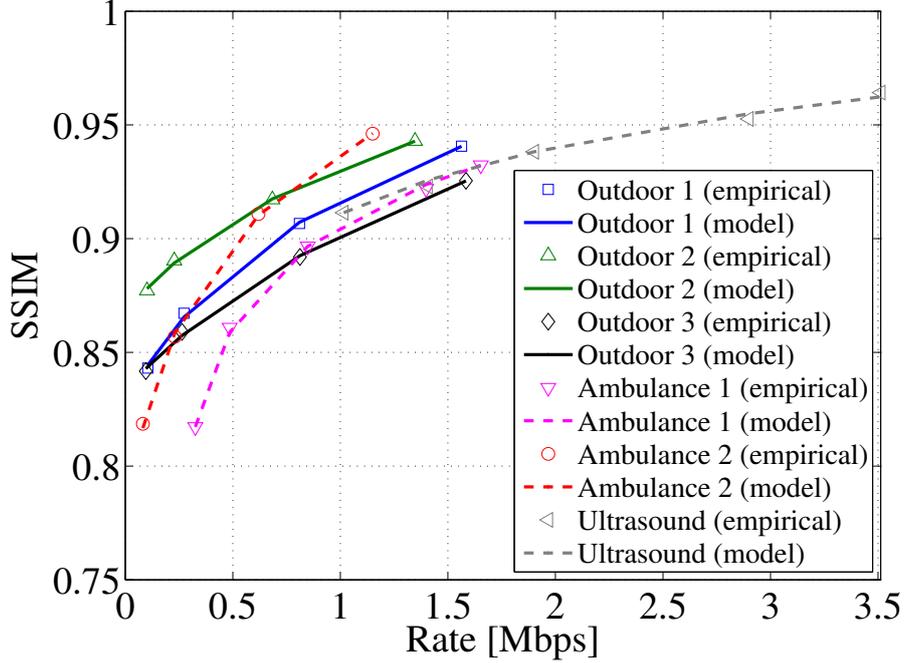


Figure 3.12: Utility models in terms of the SSIM.

utility model is considered [89]:

$$U_v(F) = a_v \ln(F - b_v) + c_v, \quad a_v > 0, b_v \leq \bar{F}_v^{bl} \quad (3.14)$$

where the utility  $U_v$  is calculated as the average SSIM of the reconstructed set of pictures. The three parameters  $a_v$ ,  $b_v$  and  $c_v$  depend on the temporal and spatial characteristics of the set of pictures  $\mathcal{I}_v$  and on the frame rate. They are derived through curve-fitting over the actual discrete empirical points. A simulation campaign has been carried out to validate the model (3.14). The results have shown almost perfect correlation with a Pearson Correlation Coefficient (PCC) always larger than 0.9999 for each set of encoded pictures. In Fig. 3.12, an example of the empirical Q-R relationship for a randomly chosen set of pictures in a IDR period is shown for five ambient video sequences and one ultrasonography video. The related Q-R curves based on model (3.14) are also reported for comparison. The encoding parameters are reported in Tab. 3.4 and 3.3. As noted in the figure, larger coding rates have been considered for the ultrasound sequence (up to 3.5 Mbps), since diagnostic videos often require very high video quality (*i.e.*, SSIM  $\geq 0.95$ ) in order to permit accurate diagnosis by remote specialists.

## Video Selection And Adaptation

It is assumed that the ambulance equipment negotiates with the LTE access network a GBR  $\bar{R}_0$  to support the emergency m-health services. Such value might be periodically updated in case of critical cell-load or bad channel conditions for the ambulance.

The m-health equipment can exploit the negotiated GBR to deliver the maximum number of videos with the highest visual quality according to different objectives and constraints. It is considered here two different strategies detailed in the following sub-sections, which jointly select the best set of videos to be transmitted and adapt the streams to the available channel. The first strategy aims at delivering the highest visual and video quality according to different priorities, derived from the CRA process, and some fairness constraints. The second strategy aims at maximizing the aggregated weighted utility, where the weights are derived from the CRA process. Specific target quality constraints for diagnostic video sequences are also considered. The two content/context-aware strategies are named fairness-based and efficiency-based optimization, respectively, and are described in the following subsections.

### Content and Fairness-Based Optimization

The proposed fairness-oriented dynamic rate adaptation strategy consists in maximizing the weighted sum of video utilities while minimizing the quality difference among the different videos. The optimization also takes into account GBR constraint, minimum utility and rate constraints, and maximum rate constraints. This kind of strategy has been first proposed in [65], addressing the problem for both empirical and semi-analytical R-D models, and then extended in [3] to consider the different priorities of the video sequences. However, the Authors in [3],[65] used the MSE as video distortion metric, without considering absolute utility constraints. Here, this approach is extended to consider the VSN and the visual quality evaluated by the CRA, with the aim to jointly select the best video set to be delivered and to provide the most fair qualities. The optimization problem is stated as follows

$$\max_{1 \leq x \leq N} \max_{F_v, v \in \mathcal{V}_x} \sum_{v \in \mathcal{V}_x} w_v U_v(F_v) \quad (3.15a)$$

$$s.t. U_v \geq U_v^{\text{target}}, \quad \forall v \in \mathcal{V}_x \quad (3.15b)$$

$$\sum_{v \in \mathcal{V}_x} \bar{H} F_v \leq \bar{R}_0, \quad (3.15c)$$

$$F_v^{bl} < F_v < F_v^{hl}, \quad \forall v \in \mathcal{V}_x \quad (3.15d)$$

$$\Delta_{\mathbf{p}}(U_i, U_j) = 0, \quad \forall i, j \in \mathcal{V}_x, \quad (3.15e)$$

where  $\bar{H}$  is the estimated overhead introduced at the different layers of the network architecture,  $F_v$  and  $U_v^{\text{target}}$ , with  $v \in \mathcal{V}_x$ , are the rate in bit per second (bps), and the target utility of video stream  $v$ , respectively.

Constraints (3.15c) and (3.15d) impose that the adapted video sequences must be supported in real-time by the available transmission GBR  $\bar{R}_0$ , thereby allocating at least the base layer rate and no more than the maximum encoding rate to each selected video stream.

The function  $\Delta_{\mathbf{p}}(U_i, U_j)$  is the utility-fairness metric for each pair of videos. It extends the fairness metric introduced in [65] to include priority information. It is defined as

$$\Delta_{\mathbf{p}}(U_i, U_j) = \begin{cases} 0 & \text{if } (i, j) \in \mathcal{F} \vee (j, i) \in \mathcal{F} \\ \delta_{\mathbf{p}}(U_i, U_j) & \text{otherwise} \end{cases} \quad (3.16)$$

with  $\delta_{\mathbf{p}}(U_i, U_j) = |p_i(1 - U_i) - p_j(1 - U_j)|$  and

$$\mathcal{F} = \{(i, j) : (U_i = U_i^{\min} \wedge p_i(1 - U_i) > p_j(1 - U_j)) \vee (U_i = U_i^{\text{hl}} \wedge p_i(1 - U_i) < p_j(1 - U_j))\} \quad (3.17)$$

where  $\mathbf{p} = [p_1, \dots, p_N] \succeq \mathbf{0}$  is a vector of positive real values, which is used to prioritize video sequences, according to some specified macro-classes of video sources, *e.g.*, to assign an higher priority to diagnostic videos with respect to ambient videos. Moreover,  $U_v^{\min} = \max(U_i^{\text{bl}}, U_v^{\text{target}})$ , and  $\wedge$  and  $\vee$  represent the logical AND and OR operators, respectively.

The function in (3.16) was introduced in [65, 90] and extends the elementary weighted fairness metric  $\delta_{\mathbf{p}}(U_i, U_j)$  to the case where the utilities  $U_i$  and  $U_j$  are constrained to their minimum and maximum values. In fact, in presence of rate and/or utility constraints, if a video achieves its maximum utility, it is reasonable to use the available resources to increase the utilities of other videos. On the other hand, in a case of scarce resources, if decreasing the rate of the  $i$ -th video is not possible since its minimum weighted utility value has been already reached, it is necessary to decrease the rate of the other videos, at the price of decreasing the related utility. As an example, by setting the value of  $p_v$  equal to 2 for the class of high priority videos, and equal to 1 for the remaining videos, the adaptation module will be able to provide at reasonable GBR values an SSIM value higher than or equal to 0.95 to the videos in the high-priority class, as long as the other videos are supported with an SSIM not lower than 0.9. It is remarkable that, as the GBR decreases, the difference in quality between macro-classes will also decrease and *vice versa*.

Finally, the weights  $w_v$  in (3.15a) are used to account for the different visual qualities according to the outcome of the CRA, and are evaluated as follows:

$$w_v = \left( \frac{q_v}{\sum_{s \in \mathcal{V}_N} q_s} \right)^\alpha, \forall v \in \mathcal{V}_x \quad (3.18)$$

where  $\alpha \geq 0$  is a parameter driving the trade-off between the number of videos to transmit and the final received quality. As it will be shown in the numerical results, a low value of  $\alpha$  favors the transmission of more videos with lower quality, while a high value of  $\alpha$  forces to select and transmit a lower number of videos with higher quality. Finally note that constraint (3.15b) enforces a target video quality (*e.g.*, when also diagnostic video are considered) in the case that a non-conservative selection of the priorities does not allow to achieve such target value.

### Problem Solutions and Algorithms

By exploiting the invertible Q-R function in (3.14), constraints (3.15b)-(3.15d) collapse in a unique constraint, *i.e.*,

$$F_v^{\min} < F_v < F_v^{hl} \quad \forall v \in \mathcal{V} \quad (3.19)$$

written in terms of minimum rate  $F_v^{\min} = \max(F_v^{bl}, F_v^{\text{target}})$ , where

$$F_v^{\text{target}} = \eta_v e^{\frac{U_v^{\text{target}}}{a_v}} + b_v \quad (3.20)$$

and  $\eta_v = e^{-\frac{c_v}{a_v}}$ .

A first step to reduce the search space of the outer discrete maximization problem considered in (3.15a), *i.e.*, the maximization with respect to  $x$ , is to exploit the feasibility condition resulting from the combination of constraints (3.19) and (3.15c), *i.e.*,

$$\sum_{v \in \mathcal{V}_x} H F_v^{\min} \leq R_0, \quad (3.21)$$

In fact, condition (3.21) allows to derive the maximum value of  $x$  that makes the problem feasible. On the other hand, the minimum value of  $x$  can be derived from the trivial condition  $\sum_{v \in \mathcal{V}_x} H F_v^{hl} > R_0$ , *i.e.*,  $x$  videos at the highest quality can be transmitted in real-time without adaptation.

Moreover, for each value of  $x$ , the inner optimization problem collapses into the following constraint satisfaction problem:

$$\sum_{v \in \mathcal{V}_x} \bar{H} F_v = \bar{R}_0, \quad (3.22a)$$

$$\Delta_{\mathbf{p}}(U_i, U_j) = 0 \quad \forall i, j \in \mathcal{V}, i > j \quad (3.22b)$$

$$F_v^{\min} < F_v < F_v^{hl} \quad \forall v \in \mathcal{V} \quad (3.22c)$$

The optimal algorithm proposed in [3] to solve a problem similar to the inner maximization in (3.22) according to a MSE-based R-D model can be now suitable extended to consider the SSIM-dependent Q-R function, by replacing the functions of (5) and (6) in [3] as follows:

---

**Algorithm 2** Pseudo code to solve problem (3.15)

---

```

1: Input:  $\mathcal{V}$ ,  $R_0$ ;  $F_v^{bl}, F_v^{hl}, U_v^{\text{target}}, w_v, \eta_v, \eta_v, a_v, b_v, \forall v \in \mathcal{V}_N$ ;
2:  $F_v^{\min} = \max(F_v^{bl}, \eta_v e^{U_v^{\text{target}}/a_v} + b_v), \forall v \in \mathcal{V}_N$ ;
3: for all  $x = N, \dots, 1$  do
4:   if  $\sum_{v \in \mathcal{V}_x} HF_v^{\min} \leq R_0$  then
5:     if  $\sum_{v \in \mathcal{V}_x} HF_v^{hl} \leq R_0$  then
6:       set  $\tilde{F}_v^{(x)} = \bar{F}_v^{hl}, \forall v \in \mathcal{V}_x$ ;
7:     else
8:        $z_v = 1, \forall v \in \mathcal{V}_x$ ;
9:       repeat
10:         $y_v = 1, \forall v \in \mathcal{V}_x; \text{cond}_h = 0$ ;
11:        repeat
12:          Compute  $\tilde{U} : \Gamma(\mathbf{y}, \mathbf{z}, x, \tilde{U}) = 0$ ;
13:           $\text{cond}_m = 0$ ;
14:          for all  $v \in \mathcal{V}_x : y_v z_v = 1$  do
15:             $\tilde{F}_v^{(x)} = \eta_v e^{\frac{\tilde{U}-1}{w_v a_v}} + b_v$ ;
16:            if  $\tilde{F}_v^{(x)} < \bar{F}_v^{\min}$  then
17:               $\tilde{F}_v^{(x)} = \bar{F}_v^{\min}; y_v = 0; \text{cond}_m = 1$ ,
18:               $y_v = 0$ ;
19:              break
20:            end if
21:          end for
22:          until  $\text{cond}_m = 0$ 
23:          for all  $v \in \mathcal{V}_x : y_v z_v = 1$  do
24:            if  $\tilde{F}_v^{(x)} > \bar{F}_v^{hl}$  then
25:               $\tilde{F}_v^{(x)} = \bar{F}_v^{hl}; z_v = 0; \text{cond}_h = 1, z_v = 0$ ;
26:            end if
27:          end for
28:          until  $\text{cond}_h = 0$ 
29:        end if
30:        if  $\sum_{v \in \mathcal{V}_x} w_v F_v^{(x)} \geq \sum_{v \in \mathcal{V}_{x+1}} w_v F_v^{(x+1)}$  then
31:           $x^* = x$ ;
32:        end if
33:      end if
34:      Transmit the video in the set  $\mathcal{V}_{x^*}$  with rate  $\tilde{F}_v^{(x^*)}$ 
35:    end for

```

---

---

**Algorithm 3** Pseudo code to solve problem (3.15)

---

```

1: Input:  $\mathcal{V}, R_0; F_v^{bl}, F_v^{hl}, U_v^{\text{target}}, w_v, \eta_v, \eta_v, a_v, b_v, \forall v \in \mathcal{V}_N;$ 
2:  $F_v^{\min} = \max(F_v^{bl}, \eta_v e^{U_v^{\text{target}}/a_v} + b_v), \forall v \in \mathcal{V}_N;$ 
3: for all  $x = N, \dots, 1$  do
4:   if  $\sum_{v \in \mathcal{V}_x} H F_v^{\min} \leq R_0$  then
5:     if  $\sum_{v \in \mathcal{V}_x} H F_v^{hl} \leq R_0$  then
6:       set  $\tilde{F}_v^{(x)} = \bar{F}_v^{hl}, \forall v \in \mathcal{V}_x;$ 
7:     else
8:        $z_v = 1, \forall v \in \mathcal{V}_x;$ 
9:       repeat
10:         $y_v = 1, \forall v \in \mathcal{V}_x; \text{cond}_h = 0;$ 
11:        repeat
12:           $\text{cond}_m = 0;$ 
13:          for all  $v \in \mathcal{V}_x : y_v z_v = 1$  do
14:             $\tilde{F}_v^{(x)} = w_v a_v \frac{\Lambda(\mathbf{y}, \mathbf{z}, x) - \sum_{n \in \mathcal{V}_x} y_n z_n b_n}{\sum_{n \in \mathcal{V}_x} w_n y_n z_n a_n} + b_v$ 
15:            if  $\tilde{F}_v^{(x)} < \bar{F}_v^{\min}$  then
16:               $\tilde{F}_v^{(x)} = \bar{F}_v^{\min}; y_v = 0; \text{cond}_m = 1;$ 
17:               $y_v = 0;$ 
18:              break
19:            end if
20:          end for
21:          until  $\text{cond}_m = 0$ 
22:          for all  $v \in \mathcal{V}_x : y_v z_v = 1$  do
23:            if  $\tilde{F}_v^{(x)} > \bar{F}_v^{hl}$  then
24:               $\tilde{F}_v^{(x)} = \bar{F}_v^{hl}; z_v = 0; \text{cond}_h = 1, z_v = 0;$ 
25:            end if
26:          end for
27:          for all  $i \in \mathcal{V}_x$  do
28:            for all  $j \in \mathcal{V}_x : j > i$  do
29:              if  $U_i(\tilde{F}_i^{(x)}) > U_j(\tilde{F}_j^{(x)}) \wedge w_i < w_j$  then
30:                 $F_j^{\min} = \eta_v e^{\frac{U(\tilde{F}_i^{(x)})}{a_v}} + b_v;$ 
31:              end if
32:            end for
33:          end for
34:          until  $\text{cond}_h = 0$ 
35:        end if
36:        if  $\sum_{v \in \mathcal{V}_x} w_v F_v^{(x)} \geq \sum_{v \in \mathcal{V}_{x+1}} w_v F_v^{(x+1)}$  then
37:           $x^* = x;$ 
38:        end if
39:      end if
40:      Transmit the video in the set  $\mathcal{V}_{x^*}$  with rate  $\tilde{F}_v^{(x^*)}$ 
41:    end for

```

---

$$\Gamma(\mathbf{y}, \mathbf{z}, x, U) = \sum_{v \in \mathcal{V}_x} \left[ y_v z_v \left( \eta_v e^{\frac{U-1}{w_v a_v}} + b_v \right) \right] - \Lambda(\mathbf{y}, \mathbf{z}, x) \quad (3.23)$$

where  $\eta_v = e^{\frac{1-c_v}{a_v}}$ ,

$$\Lambda(\mathbf{y}, \mathbf{z}, x) = \frac{\bar{R}_0}{\bar{H}} - \sum_{v \in \mathcal{V}_x} \left[ (1 - y_v) \bar{F}_v^{\min} + (1 - z_v) \bar{F}_v^{hl} \right] \quad (3.24)$$

and  $\mathbf{y} = [y_1, \dots, y_N]$ ,  $\mathbf{z} = [z_1, \dots, z_N]$  are boolean vectors that indicate whether (1) or not (0) the minimum and maximum rate constraints are active, respectively. The resulting algorithm is then iterated by varying the number  $x$  of videos to search for the best video set  $\mathcal{V}_{x^*}$  for which the weighted sum of the utility is maximum. The overall algorithm to solve the problem (3.15) in Algorithm 1 is reported, by referring the interested reader to [3] for further details.

### Content and Efficiency-Based Optimization

The proposed efficiency-oriented dynamic rate adaptation strategy consists in maximizing the weighted sum of video utilities under GBR constraint, minimum utility and rate constraints, and maximum rate constraints. With respect to the optimization in (3.15) fairness is not considered, but a video utility ordering constraint based on the visual quality is considered. The novel problem is stated as follows:

$$\max_{1 \leq x \leq N} \max_{F_v, v \in \mathcal{V}_x} \sum_{v \in \mathcal{V}_x} w_v U_v(F_v) \quad (3.25a)$$

$$s.t. \sum_{v \in \mathcal{V}_x} \bar{H} F_v \leq \bar{R}_0, \quad (3.25b)$$

$$F_v^{bl} < F_v < F_v^{hl} \quad \forall v \in \mathcal{V}_x \quad (3.25c)$$

$$(U_i - U_j)(w_i - w_j) \geq 0, \quad \forall i, j \in \mathcal{V}_x \quad (3.25d)$$

$$U_v \geq U_v^{\text{target}}, \quad \forall v \in \mathcal{V}_x \quad (3.25e)$$

where the novel constraint (3.25d) imposes that a video offering an higher visual quality must be delivered with higher video quality with respect to a video with lower ranking values. It should be reminded that the weights  $w_v$  are set according to (3.18). Problem (3.25) is a convex optimization problem. Several optimization method can be used to solve it, *e.g.*, interior-points or Lagrangian dual multipliers methods. However, constraint (3.25d) would require the derivation of  $x(x-1)/2$  dual variables, through, *e.g.*, sub-gradient method. In order to keep the complexity low, here a low-complex sub-optimal solutions is considered, which is derived by applying a method similar to the ones proposed in the previous section, *i.e.*, by

<b>Source and Coding Parameters</b>	
Number of max. video per stage	5
Frame rate	30 fps
GOP Size	8
IDR period	32
BL QP of diagnostic video	36
EL encoding	MGS
Number of EL	3
QP difference among ELs	4
<b>System model</b>	
User distribution	Uniform
Number of users	15
Number of best effort user	14
Cell layout	Single circular cell
Cell Range	300 m
<b>Channel model</b>	
Path Loss	$40 + 15.2 \log(d)$ , $d =$ distance in meter
Shadowing model	Log-normal with 4 dB standard deviation
Channel Model	ITU A extended Pedestrian model
Doppler Bandwidth	5Hz
<b>PHY model</b>	
System Bandwidth	5 MHz
Sub-carrier spacing	15 KHz
Number of sub-carriers per PRB	12
Frame duration	10 ms
Slot duration	0.5 ms
OFDM symbols per slot	7
User's power budget	23 dBm
Noise Power Density	$2 \cdot 10^{-20}$ W/Hz

Table 3.3: Simulation parameters.

Video Source	Position	Resolution	Base Layer (BL) QP
<i>Camera 1</i>	outdoor	1024x720	46
<i>Camera 2</i>	outdoor	1024x720	46
<i>Camera 3</i>	outdoor	1024x720	46
<i>Camera 4</i>	outdoor	960x540	38
<i>Camera 5</i>	outdoor	720x574	38
<i>Camera 6</i>	indoor	640x480	46
<i>Camera 7</i>	indoor	640x480	46
<i>Ultrasound</i>	indoor	640x480	38

Table 3.4: Position and resolution of the acquisition devices.

relaxing the reduced problem (3.25a)-(3.25b). In this case the solution without considering constraint (3.25d) is obtained by replacing line (15) of Algorithm 1 with the following equation:

$$\tilde{F}_v^{(x)} = w_v a_v \frac{\Lambda(\mathbf{y}, \mathbf{z}, x) - \sum_{n \in \mathcal{V}_x} y_n z_n b_n}{\sum_{n \in \mathcal{V}_x} w_n y_n z_n a_n} + b_v \quad (3.26)$$

as shown in Algorithm 2. An additional loop is consider to verify constraints (3.25d), which would lead to an optimal solution, if and only if

$$U_i(\tilde{F}_i^{(x)}) > U_j(\tilde{F}_j^{(x)}) y_i z_i y_j z_j = 1 \quad (3.27)$$

$\forall i, j > i \in \mathcal{V}_x : w_i > w_j$ , meaning that the constraint (3.25d) is violated only for videos with minimum or maximum rate constraints active. In our numerical evaluation (3.27) always holds.

### 3.4.4 Numerical Results

In the following simulations, an LTE-like access network is considered with sub-carrier spacing  $\Delta B = 15$  kHz and system bandwidth set to 5 MHz. A set of  $K = 15$  users with a maximum per-user power budget of 23 dBm is uniformly distributed in a cell of 300 m, resulting in an average SNR ranging from 5 to 28 dB. More specifically, in the proposed case study the ambulance user experiences an average SNR of 13 dB<sup>4</sup> and receives from the LTE access network a GBR  $\bar{R}_0$  with values negotiated in the range from 1 to 5.5 Mbps. All the other users are best-effort users, *i.e.*,  $K_2 = K$ . The radio channel for all users is modeled according to the ITU extended pedestrian A model, with a Doppler frequency of 5 Hz [91].

<sup>4</sup>The ambulance is assumed to be statically located on the emergency area

	Object of Interest		
	1	2	3
<i>Camera 1</i>	5790	10329	7909
<i>Camera 2</i>	4130	11413	2468
<i>Camera 3</i>	5196	4988	10304
<i>Camera 4</i>	77268	26848	0
<i>Camera 5</i>	31935	117711	0

Table 3.5: CR results ( $q_v$ ) for the three considered positions of the objects of interest.

Radio resource allocation and scheduling are performed according to the methods proposed in [4]. According to Fig. 3.9, in the first phase the ambulance sends up to  $N=5$  outdoor ambient videos, denoted as *Camera 1*, ..., *Camera 5*. In the second phase, the ambulance collects and sends 2 ambient indoor videos, namely *Camera 6* and *Camera 7*, and one diagnostic video sequence. All the videos have been acquired with a frame rate of 30fps, while their resolutions are reported in Tab. 3.4. Examples of frames acquired by the available devices are reported in Fig. 3.13.

In Tab. 3.5 the visual quality metrics defined in Sec. 3.4.2 are reported for the outdoor ambient videos available in phase 1. In particular, the  $q_v$  values corresponding to three different objects of interests within the emergency area have been listed. For example, the outcome of the CRA for the first object of interest is  $\{4, 5, 1, 3, 2\}$ , while for the third position only the first three cameras are useful, as the object of interest is outside the FoV of *Camera 4* and *Camera 5*.

SNR enhancement layers. MGS is considered to encode the quality layers. Both enhancement frames and base layer key-picture frames are used for motion compensation, thereby limiting the drift issue during the adaptation process [54]. As indicated in Tab. 3.4, the QP of the ambient videos base layer (indoor and outdoor) is set to 46, while the ultrasonography video base layer is encoded with QP equal to 38. The QP difference between adjacent enhancement layers is set to 4 for all videos. The GOP size and the IDR period are here set to 8 and to 32 frames, respectively. After encoding, the resulting quality in terms of average SSIM, ranges approximately from 0.82 to 0.95 for the ambient videos, and from 0.89 to 0.98 for the ultrasound video. The three parameters of the model in (3.14) are evaluated for each IDR period, resulting in an adaptation interval of about 1 sec. The video priority weights  $p_v$  of the ambient videos are set to 1, whereas the weight of the ultrasound video is set to 2. A target video quality for the ultrasound



Figure 3.13: Available cameras in phase 1 (*Cameras 1-5*) and phase 2 (*Cameras 6-7* and *Ultrasound*).

video is taken into account, in terms of SSIM, equal to 0.95. Finally, video play-out deadline at the receiver is set to 200 ms and the overhead factor is set to  $\bar{H} = 1$ .

The proposed strategies for joint video selection and adaptation, *i.e.*, Prioritized Equal Quality (PEQ) (see section 3.4.3) and Maximum Weighted Quality (MWQ) (see Sec. 3.4.3) are compared with an equal rate adaptation strategy, denoted with Equal rate (ER), used as benchmark. ER aims at providing fairness only in terms of assigned bit-rate, while satisfying the minimum and maximum rate constraints. It is unaware of the individual Q-R relationship of each video, *i.e.*, it is content-agnostic, and it is built according to [90]. Moreover, ER does not perform context-aware video selection.

The impact of the choice of the shaping parameter  $\alpha$ , used in the CRA-based camera weights as in (3.18) is illustrated in Fig. 3.14. Here, the weighted sum of the utilities  $\sum_{v \in \mathcal{V}_x} w_v U_v(F_v)$  (Fig. 3.14(a)), and the normalized sum of utilities  $\frac{1}{x} \sum_{v \in \mathcal{V}_x} U_v(F_v)$  (Fig. 3.14(b)) are plotted with respect to the parameter  $\alpha$  and the number of transmitted video  $x$  with a GBR  $R_0 = 3$  Mbps. The utilities are obtained as the result of the inner maximization in problem (3.15) of the PEQ strategy, *i.e.*, by considering  $x$  as an input parameter. With  $\alpha = 0$ , *i.e.*, when the visual quality is not considered in the video selection, the addition of one video generally improves the weighted sum-utility, as long as the minimum rates are supported by the GBR. This is due to the small slope of the logarithmic SSIM-to-rate relationship and to the relatively high values of the minimum SSIM of the encoded videos. A low value of  $\alpha$  is reasonable to be set up for general monitoring purposes, thereby ensuring the transmission of most ambient videos, but with low video quality. As the parameter  $\alpha$  increases, the effect of the visual quality values

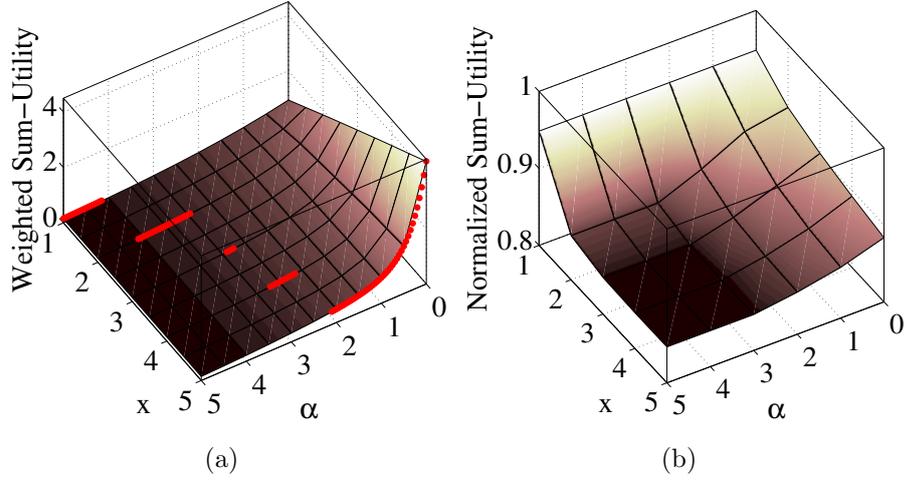


Figure 3.14: 3D plots of the normalized weighted sum of the utilities  $\sum_{v \in \mathcal{V}_x} w_v U_v(F_v)$ , and the normalized sum of the utilities  $\frac{1}{x} \sum_{v \in \mathcal{V}_x} U_v(F_v)$  with respect to the parameter  $\alpha$  and the number of transmitted video  $x$ . The utilities are obtained from the inner maximization of problem (3.15) with  $\bar{R}_0 = 3$  Mbps. The markers in (a) indicate the optimal solutions for  $x$ .

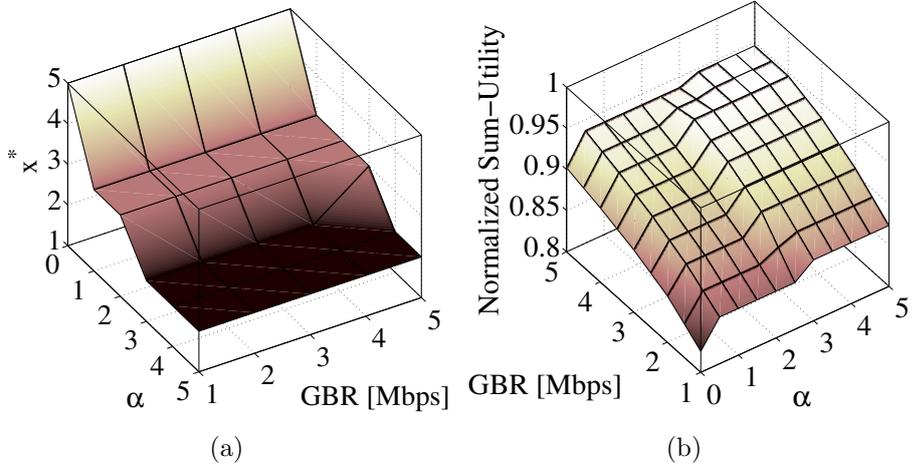


Figure 3.15: 3D plots of the optimum number of transmitted videos  $x^*$ , and the normalized sum of the utilities  $\frac{1}{x^*} \sum_{v \in \mathcal{V}_{x^*}} U_v(F_v)$  with respect to the parameter  $\alpha$  and the GBR granted to the Ambulance.

$q_v$  of the CRA is emphasized, leading to the selection of smaller sub-sets of videos with the highest visual qualities, and to a consequent improvement in the perceived video quality, as shown in Fig. 3.14(b).

Fig. 3.15 shows the impact of the available GBR on the optimum number of video to transmit, *i.e.*,  $x^*$ , (Fig. 3.15a) and on the normalized sum of the utility (Fig 3.15b), for different value of the shaping parameter  $\alpha$ . It can be noticeable that the optimum number of video  $x^*$  does not significantly depend on the GBR, and remains approximately constant when  $\alpha$  is fixed. Fig. 3.15b highlights the benefit of decreasing the parameter  $\alpha$  at low GBR, thereby significantly increasing the per-camera utility of the transmitted videos with the highest visual quality. Based on these results, in the following  $\alpha = 2.5$  is used.

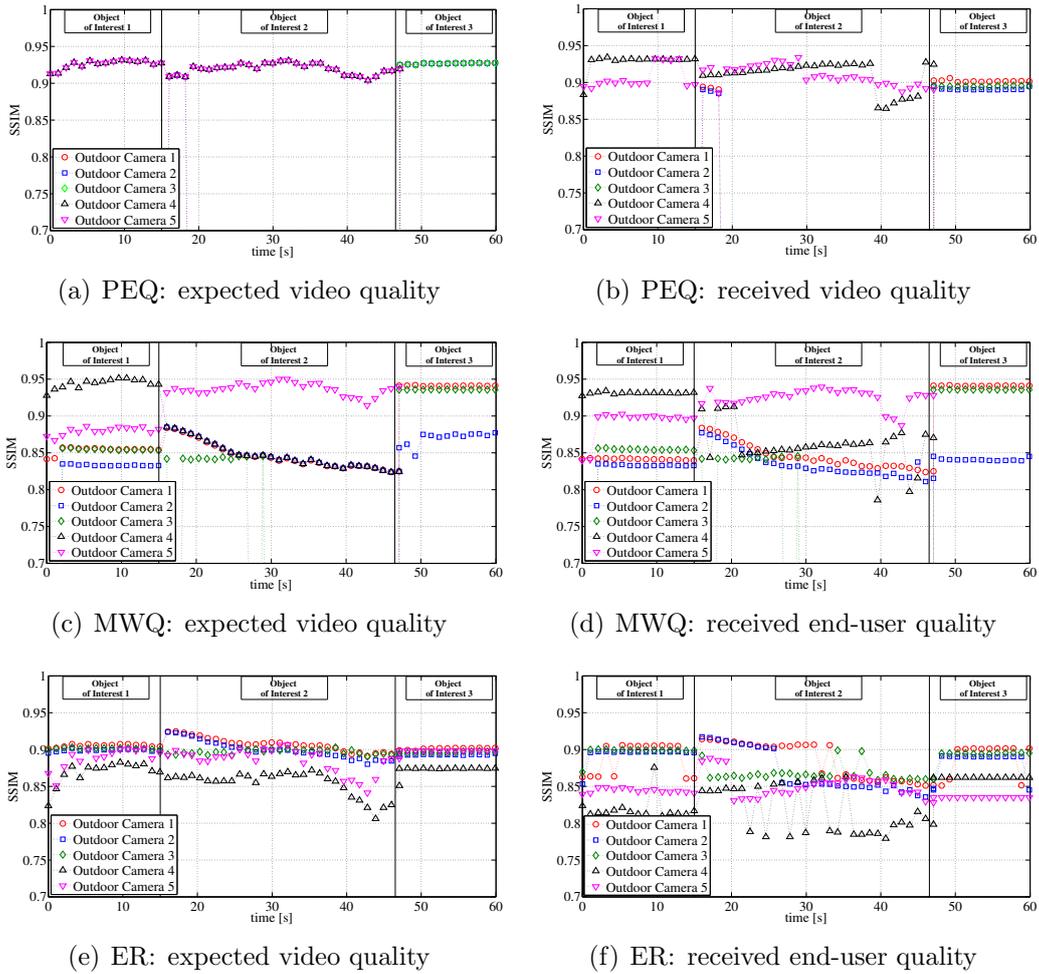


Figure 3.16: Expected and received SSIM at destination (hospital) averaged over each adaptation interval for the five outdoor video sequences transmitted in the first phase of the emergency scenario, when a GBR equal to 3 Mbps is granted to the ambulance.

Fig. 3.16 shows the expected video quality, *i.e.*, the SSIM resulting from the

continuous model in (3.14), and the received video quality at the remote hospital for the ambient videos transmitted during the first phase, according to the three different adaptation strategies. Three different object of interest are considered, whose visual weights  $q_v$  have been reported in Tab. 3.5. In this case, the GBR is set to 4 Mbps. It can be noted how the PEQ strategy optimizes the number of transmitted video by selecting the best two and the best three videos with the highest visual quality for the first two and the third object of interest, respectively, in order to jointly maximize the weighted quality and to preserve quality fairness. The difference in quality between the expected and the received quality (Fig. 3.16a vs Fig. 3.16b) is due to the gap between the continuous and the discrete utility solutions. The MWQ strategy transmits most of the videos by prioritizing the videos with the highest visual quality for each object of interest, *i.e.*, *Camera 4*, *Camera 5* and *Camera 1*, respectively. In both cases the video quality of the camera with the highest visual quality is significantly improved with respect to the ER approach, especially for the first two objects of interest. As an example, by using the ER strategy, the videos of *Camera 4* and *Camera 5* are delivered with poor quality, *i.e.*, with an SSIM close to 0.8. The improvement achievable by the proposed PEQ strategy are more evident in the second phase of the emergency situation. The results of MWQ are approximately equal to those of PEQ, due to the fact that the two ambient videos have similar spatial and temporal complexity. Therefore, in Fig. 3.17 only the PEQ results are reported. Even at large GBR, *i.e.*,  $\bar{R}_0 = 6$  Mbps (Fig. 3.17a) ER strategy may fail to provide the minimum SSIM to the ultrasound video, *e.g.*, at time instant of 8 and 32 seconds. The qualities resulting from the PEQ strategy closely follow the selected video quality priorities providing an SSIM difference that proportionally decreases as the individual SSIM increases. The minimum quality is always ensured to the diagnostic video. At lower GBR, *i.e.*,  $\bar{R}_0 = 3$  Mbps, the video quality of the diagnostic video falls below 0.9 thereby invalidating the necessary diagnostic requirement. In order to deliver the ultrasonography video sequence at the minimum required quality, the PEQ strategy discards the ambient videos, leaving all the available resources to the diagnostic video, thereby enabling effective tele-diagnosis services.

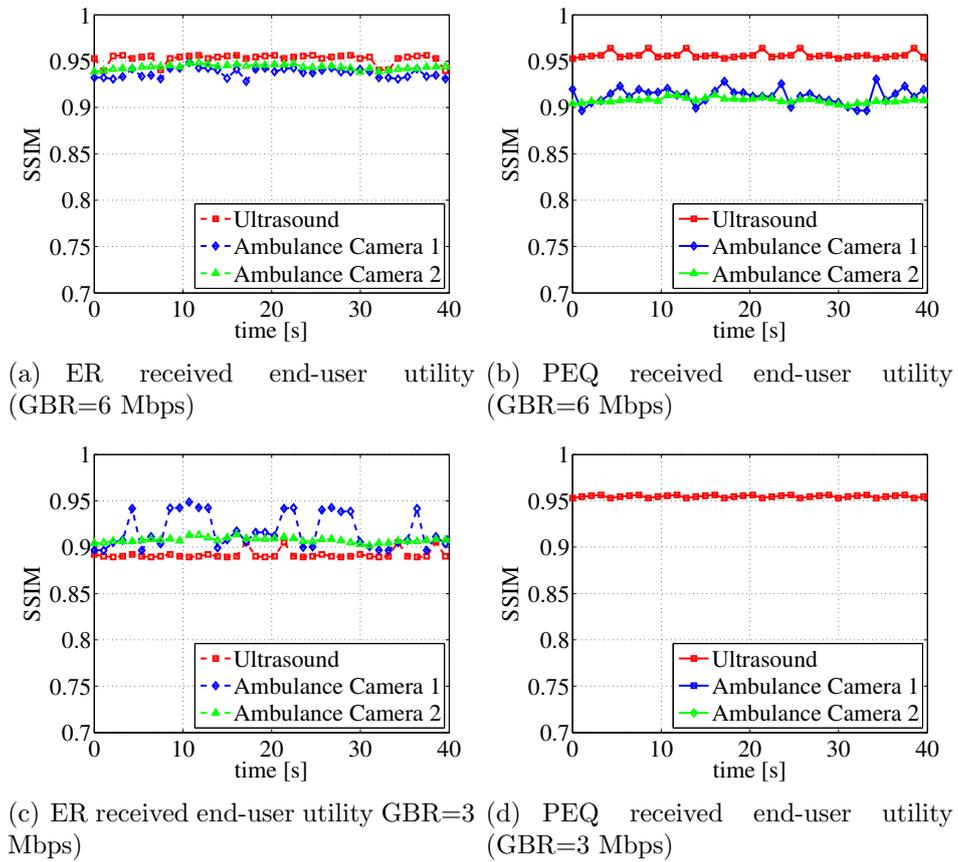


Figure 3.17: SSIM at destination (hospital) averaged over each adaptation interval for the ultrasound and the two ambient video sequences of the ambulance transmitted in the second phase of the emergency scenario, for different values of GBR granted to the ambulance. The quality weights are set to 2 for the Ultrasound video and to 1 for the ambient videos.

## 3.5 Conclusions

In this Chapter, several new SVC video adaptation solutions have been provided with the goal to satisfy the desired video and transmission quality by the above-mentioned use cases. In particular, for UC1 (Sec. 1.1), I proposed a new multiple SVC video adaptation strategy for LTE uplink transmission in mHealth emergency scenarios. The proposed solution is capable to guarantee a good end-to-end video quality despite fluctuations in the available rate and heavy traffic conditions within the LTE cell. Furthermore, several numerical results are presented showing the end-to-end video quality performance, for both diagnostic and multiple ambient videos. On the other hand, two novel content-aware aggregation and adaptation strategies and the joint application of the camera ranking algorithms (Ch. 2) are presented for UC2 (Sec. 1.2). The proposed schemes are suitable for the transmission of a set of SVC videos transmitted over LTE uplink in a mHealth emergency scenario while the context-aware camera selection algorithms can select one or more cameras taking into account specific ranking criteria mainly related to the quality of the visual representation of the object of interest. The proposed adaptation algorithm is based on modeling and evaluation of SSIM quality metric and has been tested in two operational steps. Several numerical results are presented showing the end-to-end video quality performance, for both diagnostic and ambient videos, which results to be optimal for both diagnosis and coordination purposes. Moreover, when the GBR granted to the ambulance decreases, the proposed strategy appropriately discards the videos coming from the cameras providing the lowest visual quality, according to the camera ranking results.



## Chapter 4

# Design And Implementation Of The System Simulator

The development of a simulation platform turns out to be an excellent solution for validating and also stimulating the achievement of novel results. In this regard, this Chapter is focused on the integration of the presented solutions on a common system simulator which has been developed within the CONCERTO project [14]. In particular, I focus on the design of the simulation architecture, as well as on the description of the main modules that are included in the software framework. The simulator design is strongly based on the use cases presented in Sec. 1.1 and in Sec. 1.2. The main functionalities and system requirements identified in Ch. 1 have been included in the design of the common simulation platform, in order to realistically address the described application scenarios [7]. At the same time, the simulator modules are being developed taking into account the specific research achievements of Ch. 2 and Ch. 3.

The joint development of a common simulation platform has many advantages for CONCERTO:

1. It strengthens the research cooperation among partners,
2. It permits to evaluate the system behaviour in complex scenarios, realistically taking into account many parameters affecting the communication process within different nodes of the chain and at different layers of the protocol stack,
3. It favours the identification of possible "hidden" technical problems, arising only when inserting the proposed solutions in a complete and realistic communication chain,
4. It allows a simple evaluation of very complex solutions, without requiring time and resource consuming HW/SW implementation.

The simulator, developed in the event-driven OMNET++ platform, is constituted by general macro-modules representing well-defined entities, *i.e.*, the hospitals, the emergency area that includes the ambulance and the local visual camera network, as well as the core network and the server/client terminals. Each of these macro-modules comprises one or more specific modules for which a description is schematically provided in this Chapter. Each description includes the basic input/output data information and the reference to the technical document where the modules have been detailed.

The simulator currently supports the emulation of different optimization strategies for two main use cases considered in the CONCERTO project, *i.e.*,

- "Ambulance and Emergency Area" (UC1 – see Sec. 1.1)
- "Emergency area with multiple casualties" (UC2 – see Sec. 1.2)

In the first scenario. The transmission of multiple video streams from an ambulance on the move is addressed. In particular, two ambient videos are multiplexed with an ultrasound medical stream, in order to enable remote support by specialists at the hospital. In this scenario, the impact of Packet-level Forward Error Correction (PL-FEC) on Real-Time Protocol (RTP) transmission is firstly showed from the ambulance to the CC, highlighting the quality improvements due to efficient packet loss recovery. Then, multiple communication schemes are simulated assuming different levels of optimization within the ambulance and at the eNodeB. Quality fair multiple video adaptation is compared to equal rate schemes, and the advantages of a guaranteed bit-rate policy are shown with respect to more traditional multi-user scheduling. For the use case "Emergency area with multiple causalities", the main emergency scenario addressed with the video acquisition campaign performed at the hospital of Perugia is emulated. The scenario includes two differentiated aid operational stages, *i.e.*, a first aid stage where multiple static ambient outdoor cameras acquire videos with different visual information, and an indoor phase where one injured person is loaded into the ambulance. In the latter case, two indoor ambient videos and a diagnostic ultrasound video sequence are transmitted from the ambulance. Different degrees of optimization are considered for both stages, ranging from radio resource management, *i.e.*, QoS-based optimization, to application-based enhancements, *i.e.*, QoE-aware solutions. Specifically, four strategies are evaluated, including:

- a benchmark solution where the transmitting unit at the ambulance equally divides the rate available on the wireless link among all (or to a sub-set of) the videos of the camera visual network, whereas the LTE eNodeB allocates the radio resource according to a conventional proportional fair strategy,

- a full-optimized strategy, where the ambulance divides the negotiated guaranteed bit-rate, provided by an optimized RRA at the eNodeB, among the cameras selected by a soft camera ranking algorithm, in order to provide a high quality to the diagnostic video and a lower but fair video quality to the ambient videos.

Several numerical evaluations are reported, showing the significant enhancements obtained on both the received visual and video qualities, and the large reduction of the end-to-end delay achieved by the proposed mechanisms with respect to the benchmark. The remainder of this Chapter is organized as follows: in Sec. 4.1, the peculiarities of each macro-module are initially introduced, whereas in Sec. 4.2 all significant specific modules of the simulator are schematically described. In Sec. 4.4 the detailed numerical evaluation for each use case is presented, according to the main defined Key Performance Indexs (KPIs), both QoS-based and QoE-aware [12].

## 4.1 Simulator Architecture

In order to provide realistic results, the simulation framework OMNeT++ [92] has been selected as basis to build on CONCERTO simulator. OMNeT++ is a modular, discrete time, event-driven simulation framework that provides basic modules and functionalities to build network simulators. A simulated time clock is maintained by OMNeT++ in order to avoid real time issues. One of the objectives in CONCERTO is to build on the OMNeT++ simulation framework a complete system simulator with nodes that implements a full protocol stack, from application to physical layer. Several building blocks necessary to have a complete simulator that will be used to test a subset of the use cases identified and described in Ch. 1 are provided. The simulations will provide results to evaluate the performance of the different functionalities and algorithms with anonymous medical video sequences. The achieved results will be used as feedbacks by Ch. 2 and Ch. 3 in order to improve the developed techniques. The presented simulator takes advantage of the modularity provided by OMNeT++. Several macro modules compose the simulator and each of them is, in its turn, composed by several modules and sub-modules. While macro modules represent different areas of the simulation scenario (*e.g.*, the *Hospital*, the *Emergency Area*), the modules include the different functionalities and algorithms implemented in the different nodes of the simulation. According to the different simulation scenarios the modules and sub-modules can be enabled or disabled in order to better model the corresponding use cases and to validate the different algorithms and functionalities. In Fig. 4.1 a screenshot of the CONCERTO simulator under development within the OMNeT++ Interactive Development Environment (IDE) framework is reported as an

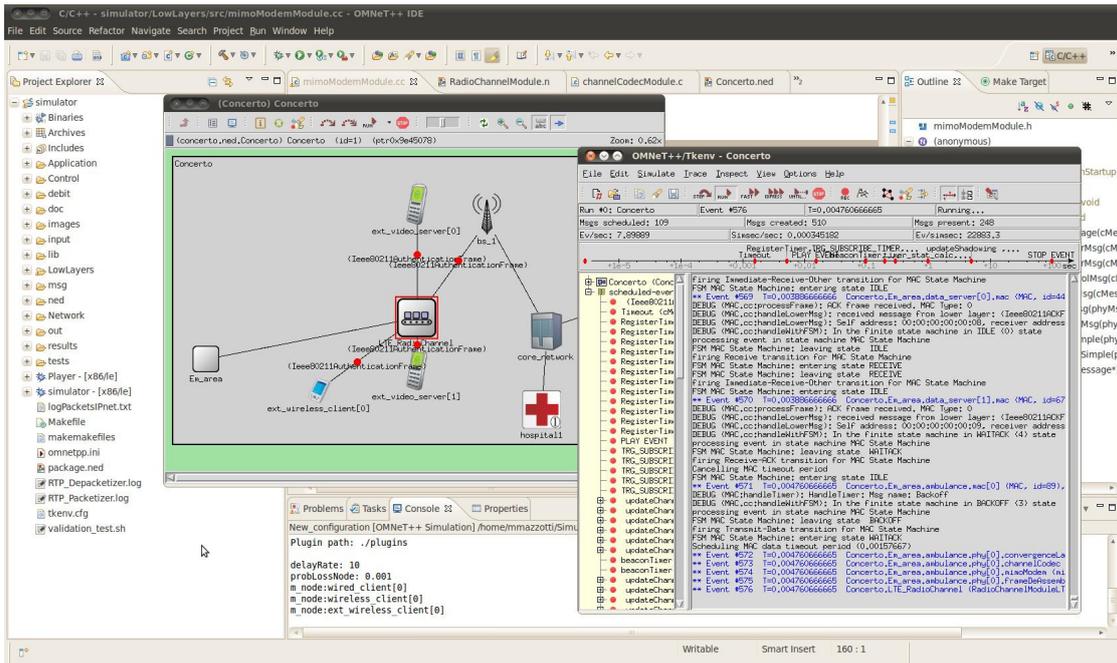


Figure 4.1: A screenshot of the OMNeT++ integrated environment adopted to develop the CONCERTO simulator.

example. In the next section a deeper view of the different macro modules in which our solution are implemented, their structure and composition is provided.

## 4.2 Macro Module Description

The general communication scenario addressed by the CONCERTO simulator is depicted in Fig. 4.2. This framework has been devised with the specific goal to evaluate some of the algorithms and solutions developed in the project in multiple healthcare and safety scenarios. In Ch. 1 two examples are described. Thanks to the modular approach followed during the simulator design, further complex scenarios, e.g. involving more terminals and information sources, may be implemented by properly replicating the basic elements reported in Fig. 4.2. As can be noted from Fig. 4.2, two hospitals/coordination centers are present, enabling the evaluation of tele-consultation and tele-diagnosis applications, in which the medical staff in one hospital (*hospital\_1*) needs assistance by specialists from another one (*hospital\_2*). The *core\_network* module simulates the behaviour of the whole Internet Protocol Version 6 (IPv6) network, constituted by multiple nodes, servers and wired links. The core network is connected to one (or more) wireless access points, represented in the simulator with the *bs\_1* base station. The radio access

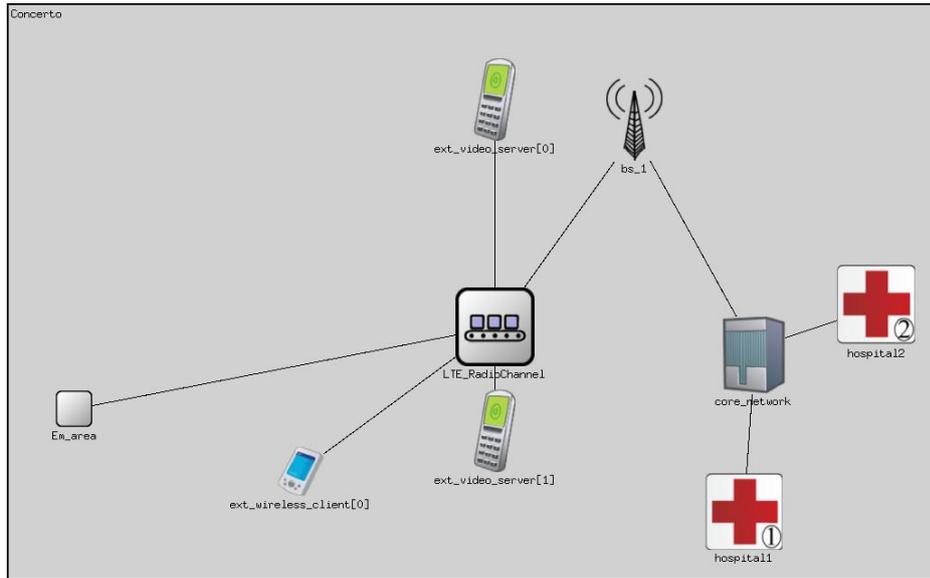


Figure 4.2: General simulation architecture.

blocks reproduce the main functionalities of 4G cellular networks at the physical and data link layers, such as multi-user scheduling, radio resource management, multiple access techniques, channel coding, modulation, etc. The 4G radio channel (in the figure indicated as *LTE\_RadioChannel*) introduces the typical impairments due to the transmission across a shared wireless medium. The most important propagation effects are taken into account, such as fast fading, shadowing, path loss, multi-user interference, etc. The 4G radio channel is shared by multiple terminals, including video servers (surveillance cameras, mobile phones, etc.) and one or more devices located in the emergency area (*Em\_area*). Depending on the tele-medicine and safety scenario considered, the emergency area can have different configurations: it can represent the ambulance rapidly moving from an accident site to the hospital (see Sec 1.1) or a complete "emergency camp" including a stationary ambulance, first-aid mobile devices and other medical equipment such as electrocardiographs, biometric sensors, ultrasound scanners, etc (see Sec. 1.2) [14]. All the basic elements reported in the Fig. 4.1 are compound modules, in their turn constituted by simpler processing blocks. This kind of scalable design of the simulator allows a clear and logical organization of the complex functionalities addressed in CONCERTO and simplifies the matching between the simulation architecture and the use cases identified in Ch. 1. In the next paragraphs the configurations of the main macro modules, in which the proposed solutions are implemented, are briefly described.

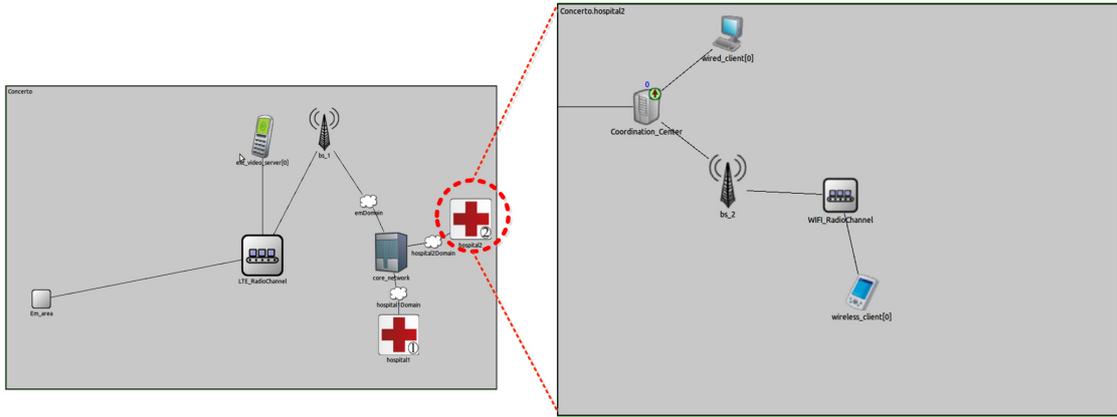


Figure 4.3: The structure of *hospital\_2* macro module.

### 4.2.1 Hospital 2

The structure of *hospital\_2* has been depicted in Fig. 4.3. A local wireless area network has been introduced, to mimic indoor mobile access through portable devices. The doctor’s mobile terminal (*wireless\_client[0]*) is connected to the core network through a WiFi access point (*bs\_2*). The transmission across the wireless channel is simulated by a radio channel module (*WIFI\_RadioChannel*). Other wired clients (*wired\_client[0]*) may also be present in the hospital. Multisource management and video ranking/prioritization algorithms are implemented in the CC as described in Sec. 1.2. Video and medical streams from the remote sources are collected by the CC, then stored and forwarded to the final user. Control signals originating from ranking, aggregation and adaptation techniques are transmitted from the hospital to the remote devices to optimize the monitoring and tele-medicine tasks.

### 4.2.2 Emergency Area

The emergency area (*Em\_area*), as depicted in Fig. 4.4, includes the ambulance, the local camera network and the medical video sources (*video\_server[i]*). Although not explicitly represented in the figure, some video sources are deployed across the entire emergency area, while others are placed on board the ambulance. The ambulance is connected to the IPv6 core network through a 4G LTE wireless link, capable to guarantee good connectivity in a wide range of mobile scenarios. Multimedia information flow from local video and medical sources to the remote hospital, passing through the aggregation server within the ambulance, while feedback and control signalling follow the reverse path, from the CC to the on-site equipment. In UC1, the ambulance is typically on the move and the different video sources are directly connected to the onboard processing unit (see Sec. 1.1

for more details), while in UC2 a stationary emergency installation is addressed, in which the medical devices can be also in proximity of the ambulance (see Sec. 1.1 for more details). In the latter case, a wireless local area network is deployed to allow short-range communication of video and data contents among the different pieces of equipment. For this reason, as depicted in Fig. 4.4, the ambulance is provided with two radio interfaces, namely IEEE WiFi (*mac[1]/phy[1]*) and 4G LTE (*lte\_mac[0]/phi[0]*). The data streams collected onboard the ambulance are multiplexed and transmitted through a single link to the LTE access network, in order to facilitate the management of the emergency communication by the 4G service provider. To coordinate the stream aggregation and transmission, a multi-source management unit is included in the ambulance processing equipment. This module is in charge of jointly selecting the target bit-rate for the multiple video encoders available on-site. The rate adaptation can be performed following different strategies, as reported in Ch. 3. For this purpose, the multi-source management units located at the CC and within the ambulance exchange several control information and feedbacks [12].

### 4.2.3 Servers And Clients

Each device acting as an information source or an information sink is implemented following the Open Systems Interconnection (ISO/OSI) protocol stack. This approach permits to evaluate the performance of novel solutions and smart algorithms keeping into account the most important features of real communication standards and their fundamental mechanisms. In Fig. 4.5, the general architecture of a video source and a client terminal has been reported. Moreover, some of the modules of Fig. 4.5 are, in their turn, compound modules constituted by simpler basic elements. For example, the *phy* module is composed by 5 sub-modules, as reported in Fig. 4.6. This section is concluded by observing that the smart functionalities addressed in the previous Chapters will be located in different modules of the communication chain. For example, the Multi-source Management (MSM) module will typically operate between the *observer\_unit* at the client side and the *controller\_video* at the streaming server. Similarly, the Application Controller will be part of the *controller\_video* within the server. On the contrary, the radio RRM module will be inserted in the *bs\_controller*, mainly operating on the MAC and PHY layers of the 4G base station (*bs\_1*). The IP content-aware cross-layer-scheduler will be distributed in the servers within the IPv6 core network, while the cross-layer signalling framework will implement novel solutions for information exchange between the *observer\_unit* at the client side and multiple elements along the communication chain [14].

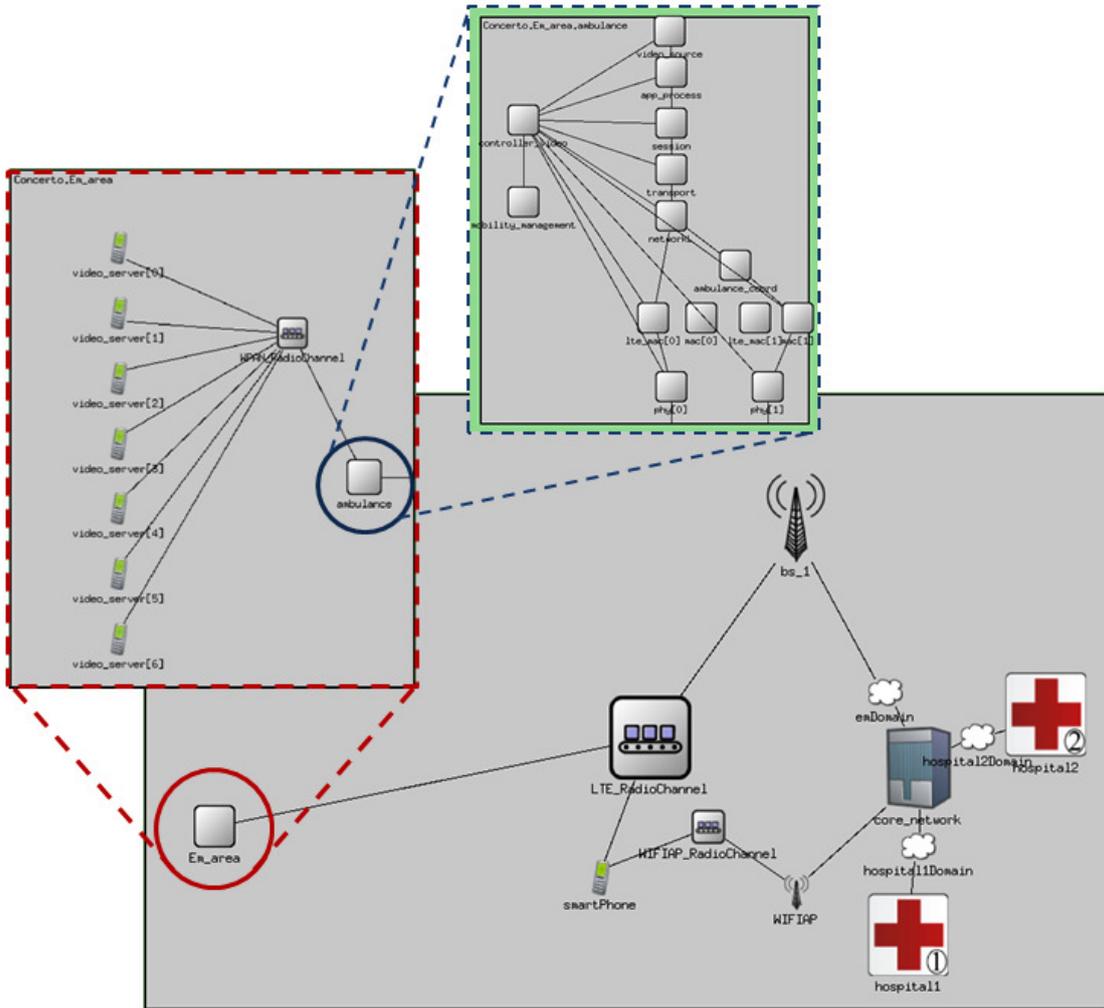


Figure 4.4: The structure of the *Em\_area* macro module.

## 4.3 Simulator Module Design

In this section, the modules included in the CONCERTO simulator are described. After a short description of the main modules' functionalities, the list of input/output information is provided, with reference to both data and control signals.

### 4.3.1 Application Controller

#### Module Description

The application controller is part of the application layer and it is the module in charge to provide the required inputs to the RTP module and to the encoder

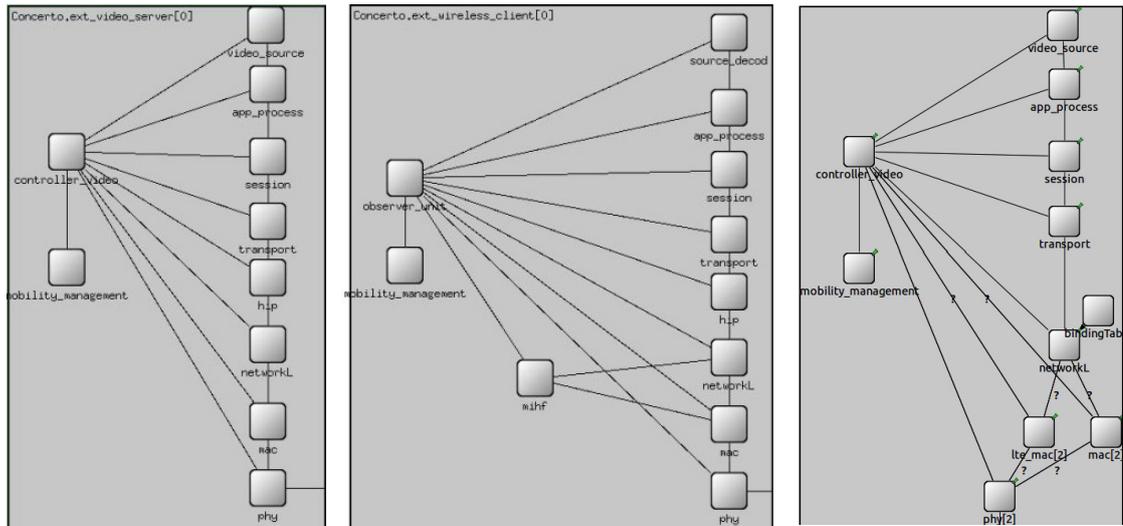


Figure 4.5: General architecture of the video sources (*e.g.*, mobile phone, smart camera, etc.), single- and multi-access client terminals (*e.g.*, remote doctor’s PC, tablet, etc.).

module (AVC). In particular, the application controller uses the expected throughput and loss probability to dynamically adapt the compression rate of the video stream to be transmitted, providing adapted frames to the RTP module. Moreover it controls the Forward Error Correction (FEC) that the RTP module will use. The application controller will be compatible with AVC and SVC streams.

### Input Data

- Data stream to transmit (from AVC/SVC encoder)
- Expected throughput (through the cross layer signaling framework)
- Expected error probability (through the cross layer signaling framework)

### Output Data

- Frames adapted to the expected throughput to RTP module
- FEC parameters to RTP module

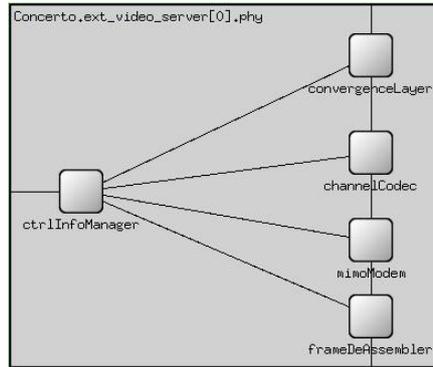


Figure 4.6: Structure of the phy layer module.

### 4.3.2 Multi-Source Management Module (Coordination Center) Module Description

The CC is developed with the aim to manage the multiple sources of information destined to end-users located in the hospital; as it is shown in Fig. 4.3, it acts as the main interface between the *Hospital 2* and the outer world. In this regard, different functionalities are implemented in this module. In particular, the CC is capable to store all the received information flows. This functionality has been implemented in order to meet the hospital internal regulations, which require that all the incoming data have to be recorded. Then, the received information are forwarded, based on specific selection criteria, to the destination user. Alternately, the stored videos can be transcoded in order to perform Transport Control Protocol (TCP)-based video communication within the hospital and compare it with the solution based on end-to-end RTP connections. The implemented video selection criteria are performed based on the video content and the video typology. For the video sources deployed across the emergency area (outdoor ambient videos in the following), the video selection is mainly based on camera ranking selection algorithms which are able to define the sub-set of cameras capable to provide best visual quality of the filmed scene in the emergency area. The implemented camera ranking strategies are mainly based on the camera pose (position and orientation) and further camera parameters as resolution and frame-rate. These information, in conjunction with the position of the point of interest, are used to define a priority policy between the different cameras. In its turn, the ambulance takes advantages of this information in order to guarantee the best quality to the prioritized selected videos during the radio resource allocation process. Again, for videos acquired directly on the ambulance, the CC provides the priority information in order to define the number of videos to be transmitted and provide the best quality to the medical data.

### **Reference For Technical Details**

The Coordination Center module functionalities are described with more details in Sec. 1.2

#### **Input Data**

- Outdoor ambient video pose (position and orientation)
- Indoor ambient video typology (ambient or medical)
- Position of the point of interest

#### **Output Data**

- Video camera priority weights

### **4.3.3 Multi-Source Management Module (Ambulance)**

#### **Module Description**

The multi-source management module in the Ambulance is in charge of determining the amount of information rate to be allocated to each video, when multiple video sources are multiplexed by the ambulance and sent through a single uplink LTE channel. This task is performed dynamically (e.g. every two seconds) according to the content characteristics of each video scene intended for transmission, thereby fulfilling the uplink transmission bit-rate provided to the ambulance equipment. The objective of this module is to provide an high quality of experience for diagnostic video sequences and a lower, but fair, quality for less critical ambient videos.

#### **Reference for Technical Details**

The multi-source management module functionalities are described with more details in Sec. 1.1

#### **Input Data**

- Set of video cameras to be transmitted
- Video camera priority weights
- Three parameters describing rate-quality model of each video scene
- Minimum and maximum allowed encoding rate

- Available uplink throughput (*i.e.*, guaranteed bit-rate if GBR-Low-rate Encoding (LRE) algorithm is applied at eNodeB)

## Output Data

- Target source bit-rate for each video

### 4.3.4 4G RRM Module

#### Module description

The 4G RRM Module is part of the "Base Station Controller" module and supports LTE-based uplink wireless transmission, specifically considered in the above-mentioned use cases. Two different RRA strategies are available: the first is a QoS-aware and channel-aware solution, based on the LRE algorithm described in Sec. 3.2.3; the second is a channel-aware proportional fairness solution used as benchmark for the final results. Both algorithms determine the power and the PRBs to be allocated to the LTE users with non-empty queue, by ensuring contiguous PRB allocation, as required by SC-FDMA technology. Open-loop LTE power control is considered to partially compensate the average attenuation. With respect to the proportional-fairness solution, the QoS-aware LRE strategy is able to differentiate among different traffic classes. Specifically, it can differentiate between two LTE macro-classes, GBR and non-GBR user flows. The resources are allocated to achieve the prescribed rate of GBR users in the short/medium term and to provide a best effort fair service to non-GBR users. In order to validate the algorithms, dummy users, in addition to the admitted users, are emulated inside this module. The dummy users can support either best-effort traffic or GBR video streaming traffic. In the latter case, realistic H.264 video sources encoded at 200 kbps are used, whereas best-effort uses are modelled as infinite buffer traffic sources. The radio resource algorithm for LTE is integrated with an AMC scheme based on the computation of the average mutual information per bit, allowing to maximize the achievable throughput while fulfilling the target requirements in terms of Bit Error Rate (BER)/BLER. The input of the AMC scheme is constituted by PRB and power allocation to the different user, while it provides the optimal CQI as output.

#### Reference for Technical Details

The 4G RRM functionalities implemented into the system simulator are described with more details in Sec. 3.1.1. The AMC scheme is based on the technique proposed in [5], extended in order to support turbo codes and uplink.

## **Input data**

The dynamic input information consists of

- Information from Internet Protocol (IP) layer of each CONCERTO user, containing
  - Amount of data in each queue
  - Head-of-line packet waiting time in each queue
- Information from PHY layer of each CONCERTO and dummy user, containing
  - Instantaneous channel gain for each sub-carrier
  - Average channel gain (path loss + shadowing)
- Allocated sub-carriers and related transmission power (AMC)
- Target BER/BLER across the LTE radio link (AMC)

Additional configuration data are:

- Bandwidth
- RRA strategy
- Number of dummy users
- Percentage of GBR-dummy users (the other are considered as best-effort)
- Rate requirements for GBR-dummy users and GBR-CONCERTO users
- Power control cell-specific parameters
- Maximum Transmission Power
- Initial GBR users load

### **4.3.5 PHY - MIMO Modem And Frame De-Assembler Modules**

#### **Module description**

The Multiple-input Multiple-output (MIMO) modem module implements the modulation and demodulation features supported by the physical layer. The modem functionalities allow to simulate the behaviour of both 4G LTE uplink and WiFi

radio links. Multiple antenna architectures are supported at both the transmitter and receiver side (typically 1, 2 or 4 antennas). For the CONCERTO simulations a Maximum Ratio Combining (MRC) scheme was selected at the receiver side. The MIMO modem module works in strict synergy with the Frame De-Assembler module, implementing OFDM modulation with CP insertion and providing the main functionalities required by OFDMA and SC-FDMA schemes. In general, the constellations supported for each sub-carrier are Binary Phase Shift Keying (BPSK), QPSK, 8PSK, 16QAM and 64QAM. When LTE links are considered, only QPSK, 16QAM and 64QAM (for the downlink) are possible. Moreover, single antenna LTE terminals and 2-antennas eNodeB were assumed. The MIMO modem module constitutes a key element for the AMC solution for LTE (see the description of the 4G RRM module).

### **Reference for Technical Details**

Standard OFDM-based modulation is implemented, supporting different types of multiple access to the shared channel (WiFi and LTE).

### **Input data**

The input information consist of configuration information and data packets to be processed:

- Modem type
- Number of antennas
- Constellation size (BPSK, QPSK, 8PSK, 16QAM, 64QAM)
- CP duration
- Number of sub-carriers
- Transmission bandwidth
- Resource block allocation information
- (Max) transmission power
- Antenna gain and cable loss
- Rx noise figure
- Data packets from the Channel Codec module (Modulator)
- Received symbols (Demodulator)

## Output data

The output information consist of processed symbols and feedback information:

- Modulated symbols (Modulator)
- Log-Likelihood Ratio (LLR) associated to the received symbols (Demodulator)
- CSI, in terms of sub-carrier gains and signal-to-noise ratio (Demodulator)

## 4.3.6 Radio Channel Module

### Module description

The Radio Channel module simulate the effects of packet transmission across radio links. Several types of wireless radio channel have been realized and integrated in the simulator:

1. LTE channel
  - a) based on the ITU models:
    - i ITU Pedestrian and Vehicular A,
    - ii ITU Pedestrian and Vehicular B,
    - iii ITU Extended Pedestrian A,
  - b) supported bandwidth: 1.4MHz, 3MHz, 5MHz, 10MHz, 15MHz, 20MHz
2. WiFi channel, based on the ITU models:
  - a) ITU Residential,
  - b) ITU Office
  - c) ITU Commercial.
3. Uncorrelated Block Fading (UBF) channel
  - a) Fully configurable

The implemented solution allows to simulate several kinds of OFDM signals, like those supported by LTE and WiFi systems. Possible presence of multiple transmitting and receiving antennas is considered. In addition to path loss depending on the terminal position, Rayleigh distributed channel gains and log-normal shadowing model the fading effects due to time-varying multipath propagation and obstacles, respectively. Fast fading samples are derived based on the selected

channel model and on the mobility of the terminals, defined by their velocity and motion direction. Slow fading samples are calculated taking into account their temporal correlation, based on the user equipment velocity. This model permits to achieve a good trade-off between simulation realism and computational burden. In order to lighten the simulation complexity and, consequently, to reduce the simulation time, simplified versions of this module have been implemented and can be adopted in specific parts of the simulator. For example, the radio channel modelling the local/personal links between the cameras and the medical devices within the emergency area will be simulated assuming simple packet erasure channels with a predefined loss probability[12].

### **Reference for Technical Details**

Wideband radio channels have been implemented according to the ITU models described in [93, 94].

### **Input data**

The input information consist of configuration information and modulated symbols to be processed:

- Channel type and model (LTE, WiFi, UBF)
- Transmission bandwidth
- Terminal position, velocity and direction
- Time and frequency block length (with UBF)
- Number of carriers
- Number of TX and RX antennas
- Log-normal fading parameters
- Transmitted symbols

### **Output data**

The output information consists of processed symbols and feedback information:

- Received symbols
- Channel state information

## 4.4 Simulation Results

### 4.4.1 KPI Evaluated Through Simulations

#### End-to-end Quality

The end-to-end video quality is evaluated through multiple objective metrics, namely PSNR and SSIM. This choice allows to validate the numerical results reliably and, at the same time, to compare them easily with other future solutions. In addition to the quality metrics, the QoE is evaluated through the end-to-end delivery delay. Low latency enables real-time or quasi real-time services, characterized by the possibility for the remote specialist to provide an interactive support to the on-site staff.

#### Transmission Quality

Radio link quality is evaluated by means of traditional metrics, such as achieved throughput and link delay. In particular, in the considered use cases the focus is on the performance across the LTE mobile radio access, as it constitutes the most critical wireless link.

### 4.4.2 Simulated Scenarios

#### UC1: Ambulance And Emergency Area

This scenario mainly focuses on multimedia communication between the ambulance on the move and *hospital\_2* where a doctor can utilise multimedia content from the ambulance for different purposes such as monitoring, coordination and/or consultation. The main video and medical data streams are originated by pieces of equipment on the ambulance and are received by a client at the hospital. At the hospital a doctor, using these video and data, can monitor the patient condition and coordinate the first-aid operations. Some medical devices may be connected via radio with the ambulance equipment. Since the ambulance is typically moving, sometimes also at high speed, a 4G link is supposed to be available to communicate with the hospital. On the reverse link, from the hospital to the ambulance, multiple feedbacks and signalling information are transmitted. Based on the automatic monitoring systems or user-specified preferences, proper coordination signalling can help to improve the on-board assistance and the degree of remote context-awareness, increasing the efficiency and effectiveness of the interventions on the patient. To this purpose, the first-aid staff on the ambulance may also have real-time access to remote databases containing the most significant medical records about the patient's health history. The simulated scenario is described in Sec. 1.1.

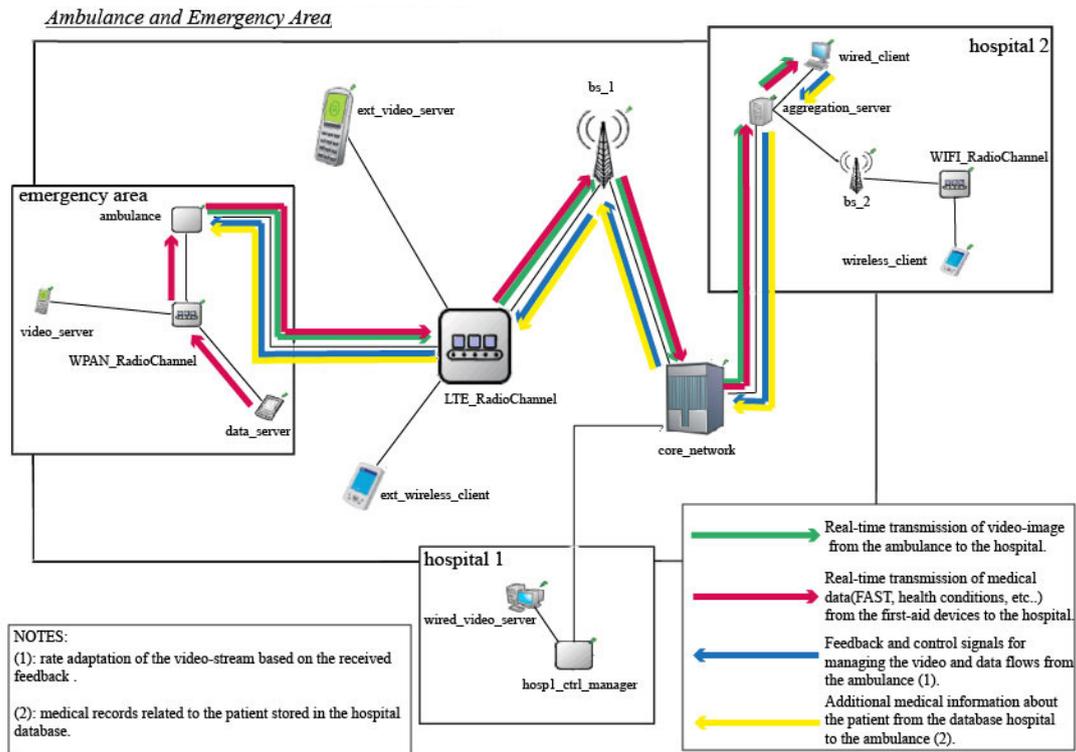


Figure 4.7: Ambulance and emergency area: information flows.

The simulation is focused on transmission of two ambient video streams and one ultrasound sequence acquired inside the ambulance. In this way, the medical assistance is simulated and performed after loading the patient on the vehicle with the support of advanced tele-diagnosis services. The ambulance is assumed to be moving (most likely fast) towards the hospital. The main characteristics of the options implemented in the CONCERTO simulator and evaluated in this user scenario are summarized in the following.

- **Ambulance** – The processing unit on-board the ambulance is responsible to collect and aggregate the multimedia streams from two ambient cameras and from one medical device available in the ambulance and to transmit them through the LTE link, toward the CC/hospital. To perform its processing, the ambulance exploits the ranking information received from the CC and the achievable throughput across the LTE link. The target video coding rates of the selected sources are jointly adapted to guarantee the optimal level of quality (in terms of PSNR and SSIM) compatible with the available radio resources and communicated to the application controller of each encoder.

Two techniques have been compared to perform rate adaptation:

- ER, in which the available throughput is equally divided by the number of sources selected for transmission, without taking into account the final video quality.
- Quality fair (QF), in which the target source rates are jointly determined splitting the overall throughput not uniformly, but aiming at achieving the same video quality for two ambient streams. On the contrary, for the medical stream a higher quality is selected, in order to allow tele-diagnosis support.

Together with the rate adaptation, the efficiency of technique PL-FEC based on non-binary LDPC codes have been evaluated in this scenario. Three cases have been evaluated:

- PL-FEC is disabled with ER adaptation technique (benchmark)
  - PL-FEC is enabled with ER adaptation technique
  - PL-FEC is enabled with QF adaptation technique
- **LTE eNodeB** –Two scheduling and radio resource allocation techniques have been implemented in the simulator: a traditional Proportional Fair (PF) approach and the GBR-Adaptive Low-rate Encoding (ALRE) solution described in [11]. In the latter case, the ambulance is treated by the LTE access network as a privileged user, and a guaranteed bit-rate is provided by the system based on a low-complexity processing scheme fully compatible with the LTE uplink standard. The simulator takes into account the presence of additional non-CONCERTO users in the cell requiring bandwidth for their uplink transmissions. These techniques have been evaluated with the three cases of optimization techniques described above.

## Simulation Parameters

In this paragraph, the main simulation settings used for UC1 are reported, with reference to different sections of the communication chain.

### *Video Sources:*

The initial encoding rate for each camera has been set to 500 kbps. The rate adaptation period has been set equal to the GOP size, namely 1s (25 frames). More details are presented in Tab. 4.1.

Use Case 1		
Medical Video ( <i>video_server[4]</i> ) 640x480, 25fps		
Ambient Camera 5 ( <i>video_server[5]</i> ) 640x480, 25fps		
Ambient Camera 6 ( <i>video_server[6]</i> ) 640x480, 25fps		

Table 4.1: The used video source for simulating the UC1.

Emergency Area	
Type of radio link	IEEE 802.11g
Channel model	ITU Commercial Area
Modulation	OFDM (QPSK on each subcarrier)
PHY FEC	Punctured convolutional code, $R_c=1/2$
Cyclic Prefix	1/16

Table 4.2: EA parameters for simulating the UC1.

## UC1: Numerical Results

The simulation campaign realized to evaluate UC1 consists in a series of tests taking into account different combinations of optimization techniques developed in the project focusing especially on PL-FEC. This system benchmark is compared with different rate adaptation techniques (ER and QF) at the ambulance with PL-FEC enabled or disabled, different resource allocation techniques at LTE eNodeB (PF and GBR-ALRE). The set of performed simulations is reported in Tab. 4.5 where the Sim.UCI.00 case corresponds to the system benchmark, representing what may be achieved nowadays with traditional technologies. The performance evaluation of the proposed PL-FEC schemes is proposed. As it is shown in Tab. 4.5, in the benchmark case the PL-FEC module is deactivated while in Sim.UCI.00.RTP and Sim.UCI.11.RTP it is operating in order to prevent packet drops during the wireless transmission and losses during the IP routing and delivery. Fig. 4.8,4.9,4.10,4.11 and Fig. 4.12,4.13,4.14 show the received video quality, in terms of PSNR and SSIM for the RTP-based simulations. In the bottom right sub-figures, the mean PSNR and SSIM value is depicted for each user. In the benchmark case, it is

Hospital Area	
Type of radio link	IEEE 802.11g
Channel model	ITU Office Area
Modulation	OFDM (QPSK on each subcarrier)
PHY FEC	Punctured convolutional code, Rc=1/2
Doctor's UE velocity	3 km/h
Distance to the AP	Between 1.5m and 30m
Shadowing Sigma dB	Between 4dB and 6dB

Table 4.3: *Hospital\_2* parameters for simulating the UC1.

noticeable how, in presence of transmission errors and packet dropping, the final quality can be dramatically compromised. On the contrary, the application of a PL-FEC scheme based on a non-binary LDPC code developed in [11] allows to reach satisfying level of quality recovering the packet losses. However, in Sim.UCI.00.RTP the medical video does not obtain a sufficient level of quality to be used for medical purposes such as tele-diagnosis. Minimum tolerable values for diagnostic evaluations are 35 dB of PSNR and 0.95 of SSIM. Finally, the combination of the PL-FEC strategy and the proposed quality fairness and GBR-ALRE criteria provides the suitable protection from packet losses and at the same time guarantees higher level of quality for the medical user and fair quality for the remaining ambient videos.

## UC2: Emergency Area With Multiple Casualties

In Fig. 4.15 the main information flows are reported between the EA and *hospital\_2*. Within the emergency area, local or personal area links allow the short-range communication between ambulance, cameras and on-site medical devices. In this scenario, the ambulance is typically stationary and acts as an information hub for all the local sources of information. A 4G link (*e.g.*, LTE/LTE-A) allows a bidirectional connection with remote doctors and crisis coordinators. Besides, those directly belonging to the emergency area, other cameras may be present (*e.g.*, mobile phones of bystanders, surveillance cameras, etc.), transmitting useful information to the emergency coordination center at the hospital. As a result, multiple video and biomedical data streams are transmitted from the emergency area to the hospital, across multiple wireless and wired links. A professional within *hospital\_2* receives the information streams through a WiFi link and orchestrates the operations from remote. Automatic or user-driven management systems coordinate the adaptation, information fusion, protection and prioritization tasks by transmitting proper feedback and control signalling [14]. The simulated emergency area is described in Sec. 1.2. The simulated emergency situation is structured according to Fig. 4.16. In the first phase, the first aid staff assist two injured people on the ground, located some distance apart. Three points of interests are considered in se-

4G Link	
Type of radio link	LTE Uplink
Duplexing Mode	FDD
Modulation	OFDM (adaptive constellation based on AMC)
Multiple Access Type	SC-FDMA
Minimum Allocable Resource Block	1 PRB (=12 subcarriers x 7 OFDM symbols)
PHY FEC	Turbo code (adaptive Rc based on AMC)
Cyclic Prefix	4.69us (short CP)
Adaptive Modulation and Coding (AMC)	Enabled
Radio Resource Management	PF / GBR-ALRE
Ambulance (Tx)	
<i>Channel model</i>	ITU Vehicular A
<i>Velocity</i>	50 km/h
<i>Max TX Power</i>	18dBm
<i>Antenna Gain</i>	2.0dBi
<i>Nr. of Antennas</i>	1
<i>Lognormal Shadowing sigma dB</i>	4dB
<i>Initial distance from the eNodeB</i>	500m
eNodeB (Rx)	
<i>Antenna Gain</i>	18.0dBi
<i>Nr. of Antennas</i>	2
<i>Cable Loss</i>	2.0dB
<i>Rx Noise Figure</i>	2.0dB
Additional Non Concerto Users (Tx)	
<i>Nr. of Non Concerto Users</i>	10
<i>Channel model</i>	ITU Pedestrian A
<i>Distance to the eNodeB</i>	Between 50m and 1500m

Table 4.4: eNodeB parameters for simulating the UC1.

Simulations	PL-FEC	Ambulance	LTE eNodeB
Sim.UCI.00	off	ER	PF
Sim.UCI.00.RTP	on	ER	PF
Sim.UCI.11.RTP	on	ER	GBR ALRE

Table 4.5: Set of Simulations of UC1.

quence, corresponding to the two on-the-ground positions and to the loading area behind the ambulance, respectively. Four ambient cameras are deployed in the area and available to acquire and transmit information. Several simulations have been done considering different sets of transmitted videos: first it is assumed that the entire set of video streams has to be sent to the CC, then the evaluation of the case in which just two videos sources out of the available four are selected through fixed or dynamic mechanisms and transmitted to the remote user is accomplished. The second simulation phase consists in the transmission of two ambient video streams and one ultrasound sequence acquired inside the ambulance. In this way,

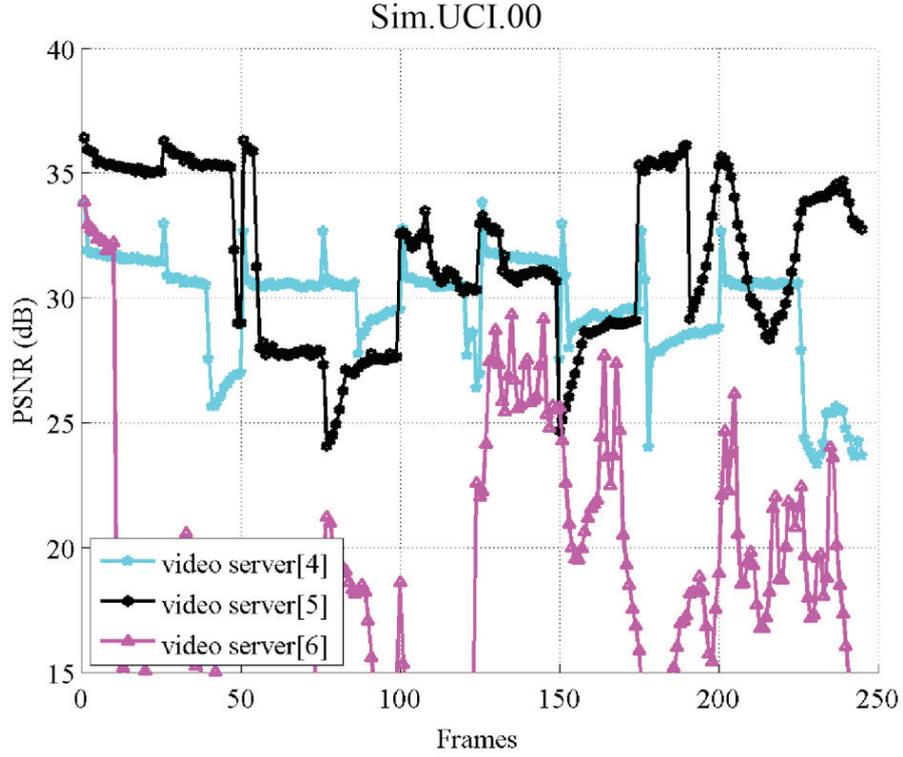


Figure 4.8: End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

the medical simulated assistance is performed after loading the patient on the vehicle with the support of advanced tele-diagnosis services. Differently from UC1, here the ambulance is assumed as stationary. The CONCERTO modules responsible to coordinate the communications between the emergency area and the remote hospital are depicted in Fig. 4.17. The position of each is also reported in the figure, in order to facilitate the mapping of the processing units within the general architecture. For each element, different optimization techniques are evaluated, allowing a performance comparison between the solutions developed in the project and traditional schemes. The main characteristics of the options implemented in the CONCERTO simulator are summarized in the following.

- **Coordination Center (CC)** When Source Ranking (SR), the CC transmits a prioritized list of video sources to the ambulance adaptation unit, in charge of enabling and adapting all the cameras in the emergency area. The ranking techniques developed in this work are described in Ch. 2. In particular, when only two videos are selected for transmission, ranking information are used to dynamically switch among the transmitted cameras, based on the location

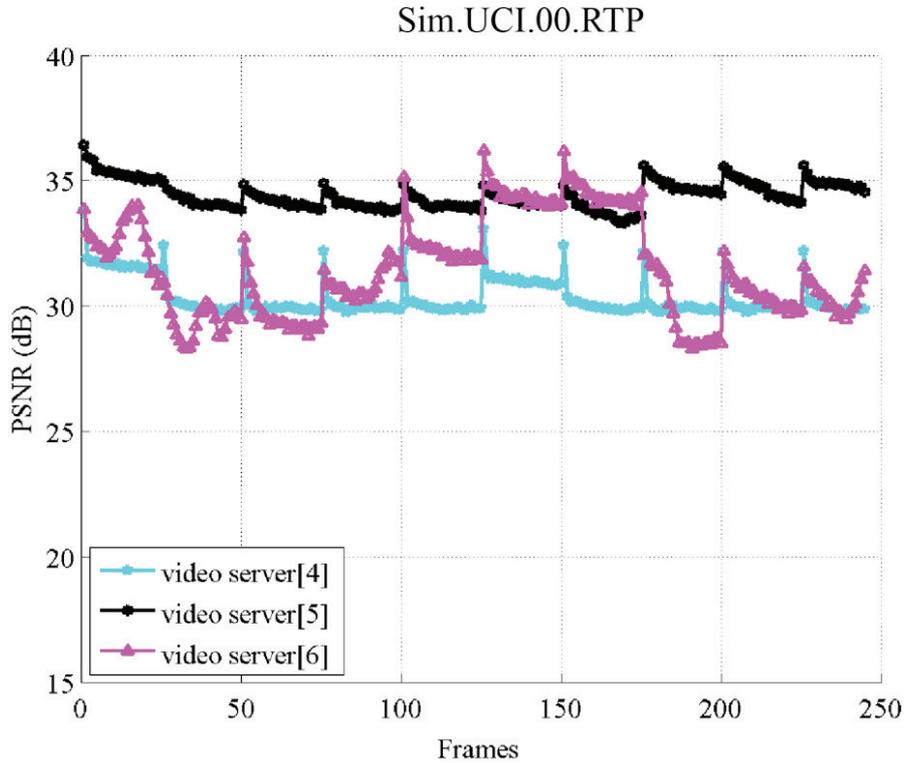


Figure 4.9: End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

of the current point of interest. The hard ranking information calculated for Phase 1 of UC2 are reported in Tab. 4.6:

Phase 1		
Point of Interest nr. 1	Point of Interest nr. 2	Point of Interest nr. 3
1) Camera 3 ( <i>video_server[2]</i> )	1) Camera 4 ( <i>video_server[3]</i> )	1) Camera 0 ( <i>video_server[0]</i> )
2) Camera 4 ( <i>video_server[3]</i> )	2) Camera 3 ( <i>video_server[2]</i> )	2) Camera 1 ( <i>video_server[1]</i> )
3) Camera 0 ( <i>video_server[0]</i> )	3) Camera 1 ( <i>video_server[1]</i> )	3) Camera 3 ( <i>video_server[2]</i> )
4) Camera 1 ( <i>video_server[1]</i> )	4) Camera 0 ( <i>video_server[0]</i> )	4) Camera 4 ( <i>video_server[3]</i> )

Table 4.6: The cameras listed in red are selected for transmission from the emergency area when the source ranking is activated in Phase 1.

- **LTE eNodeB** – Two scheduling and radio resource allocation techniques have been implemented in the simulator: a traditional PF approach and the GBR-ALRE solution described in Ch. 3. In the latter case, the ambulance is treated by the 4G access network as a privileged user, and a guaranteed

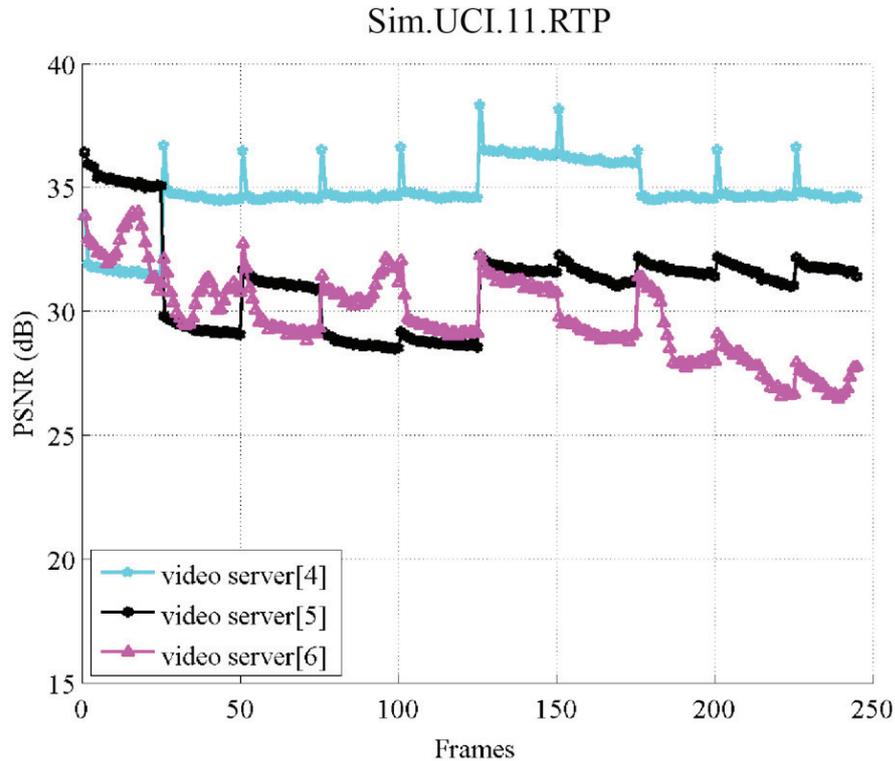


Figure 4.10: End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

bit-rate is provided by the system based on a low-complexity processing scheme fully compatible with the LTE uplink standard. The simulator takes into account the presence of additional non-CONCERTO users in the cell requiring bandwidth for their uplink transmissions.

- Ambulance** – The processing unit on-board the ambulance is responsible to collect and aggregate the multimedia streams from all the cameras available in the emergency area and to transmit them through the LTE link, toward the CC/hospital. To perform its processing, the ambulance exploits the ranking information received from the CC and the achievable throughput across the LTE link. The target video coding rates of the selected sources is jointly adapted to guarantee the optimal level of quality (in terms of SSIM) compatible with the available radio resources and communicated to the application controller of each encoder. Two techniques have been compared through simulation to perform rate adaptation:
  - ER, in which the available throughput is equally divided by the number

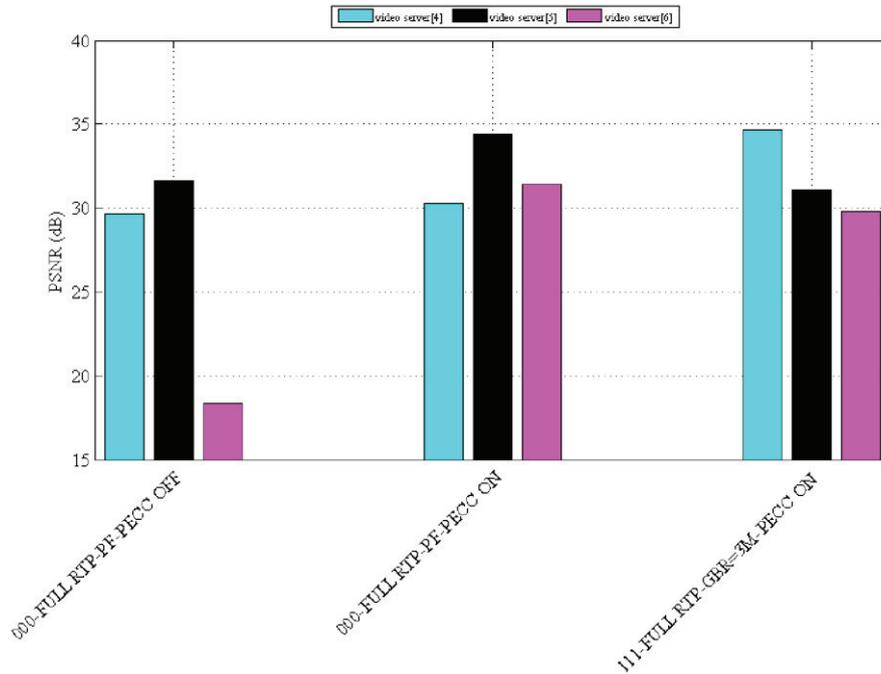


Figure 4.11: End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

of sources selected for transmission, without taking into account the final video quality.

- QF, in which the target source rates are jointly determined splitting the overall throughput not uniformly, but aiming at achieving the same video quality for all ambient streams. On the contrary, for the medical stream an higher quality is selected, in order to allow tele-diagnosis support.

## Simulation Parameters

In this paragraph, the main simulation settings used for UC2 are reported, with reference to different sections of the communication chain.

### *Video Sources:*

The initial encoding rate for each camera has been set to 1 Mbps. The rate adaptation period has been set equal to the GOP size, namely 1s (25 frames). More details are presented in Tab. 4.7 and in Tab. 4.8.

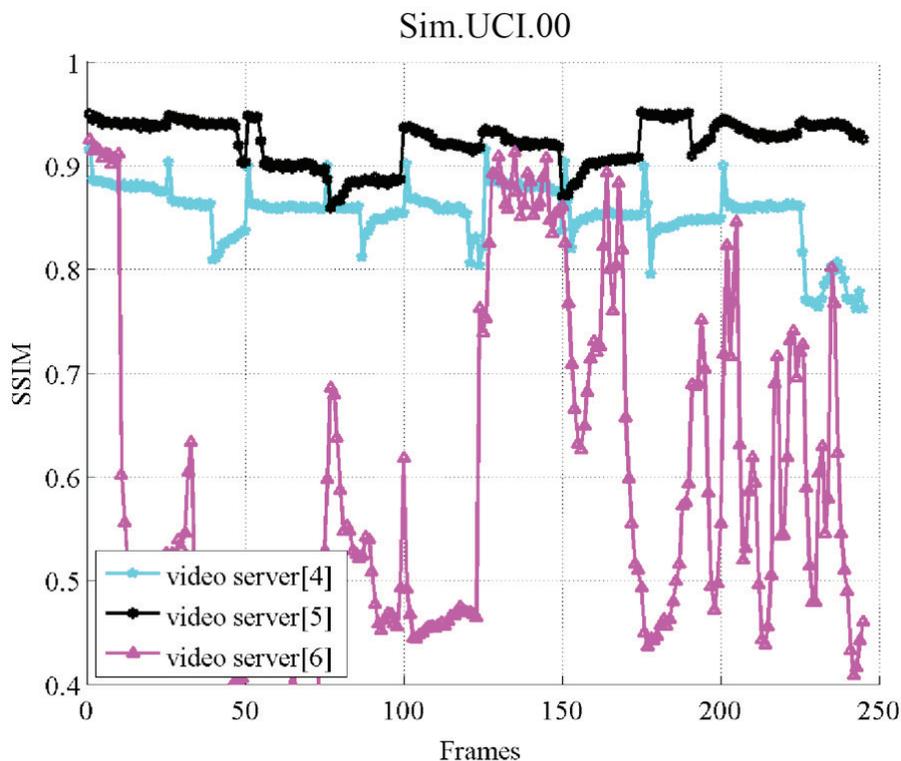


Figure 4.12: End-to-end SSIM for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

## UC2: Numerical Results

The simulation campaign realized to evaluate UC2 consists in a series of tests taking into account different combinations of optimization techniques developed in the project. In Tab. 4.12, the entire set of performed simulations is reported, showing how different CONCERTO solutions can be combined to obtain different trade-offs between performance and complexity. In particular, the Sim.UCI.000 case corresponds to the system benchmark, representing what may be achieved nowadays with traditional technologies. At the opposite side, Sim.UCI.111 represents a fully optimized CONCERTO solution[12]. The scenarios involving GBR-ALRE optimization in the LTE uplink requires that the 4G service providers implement some of the resource allocation and scheduling solutions developed in the project and described in Ch. 2 and in Ch. 3. On the contrary, the optimization processing performed at the CC and within the ambulance depends only on the emergency management organization, and relies on exchange of control and feedback information. The entire set of simulations reported in Tab. 4.12 has been realized assuming no transmission errors across the radio links and no packet losses within

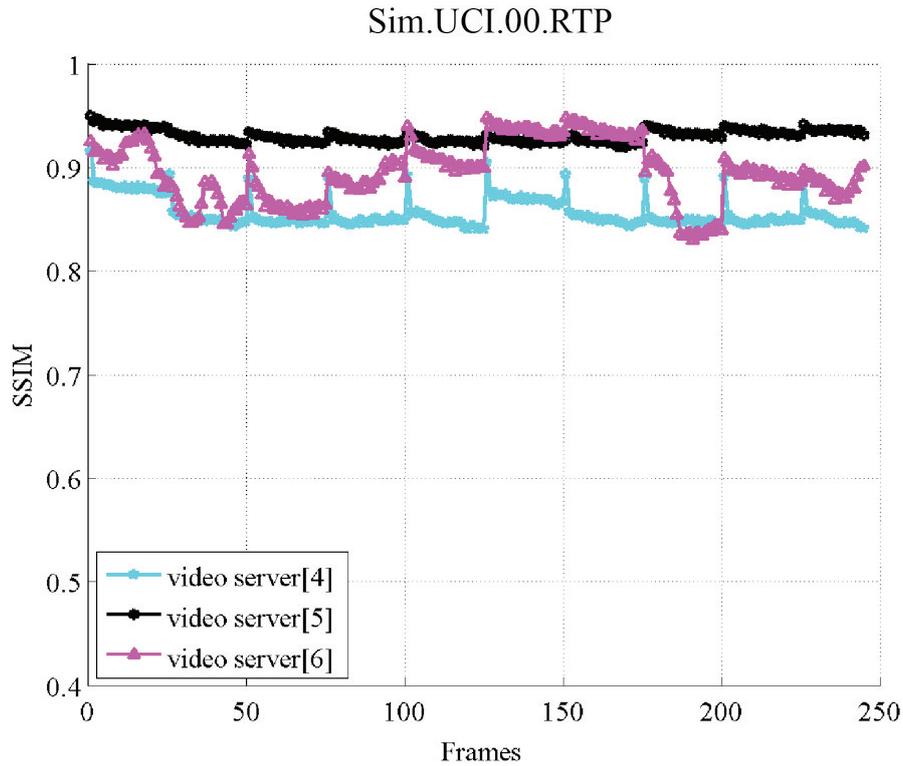
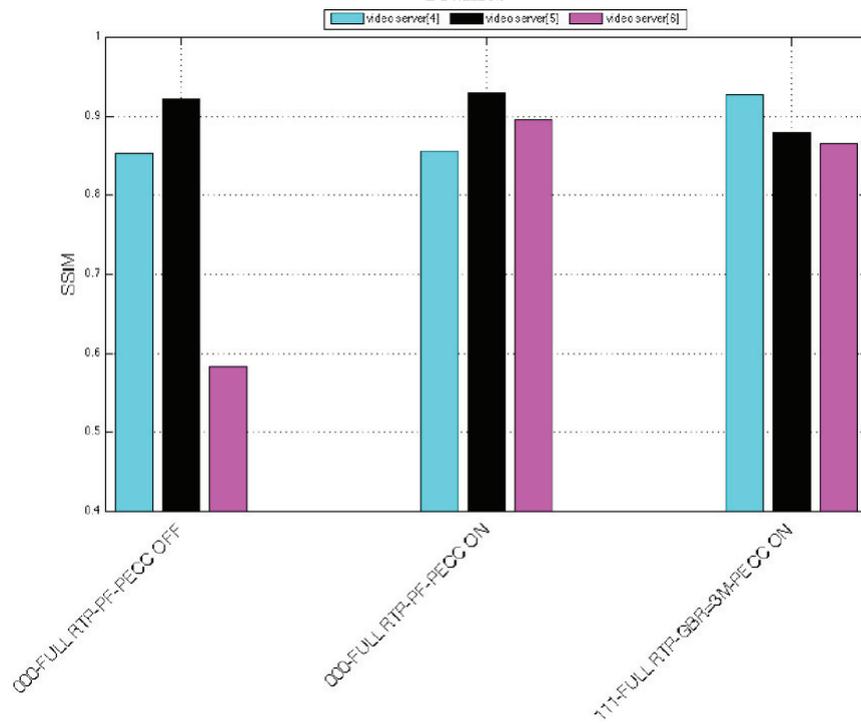
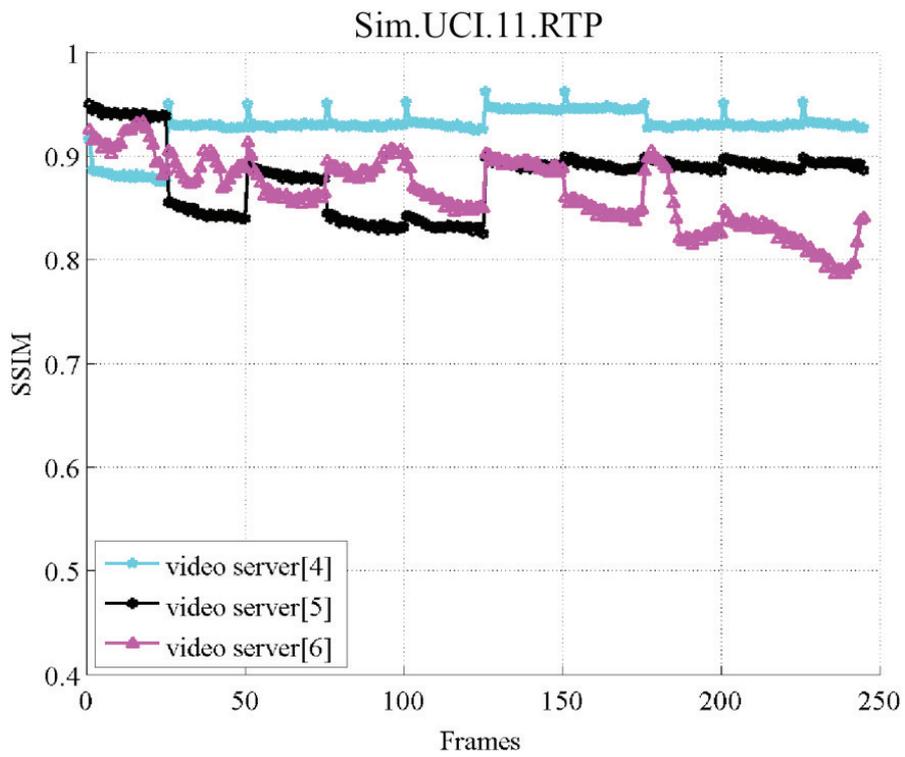


Figure 4.13: End-to-end SSIM for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection.

the IPv6 core network. However, losses are possible due to buffer overflow at the radio link MAC layer. In the next paragraphs, the performance of the video selection and adaptation solutions is presented, as well as the LTE radio resource allocation under these conditions. In the following, several curves are reported for each scenario, depicting the performance in terms of video quality, end-to-end latency, achieved throughput/delay across the LTE radio link and source coding rates. Some observations are also reported, while a comparison of the different optimization strategies considered for UC2 is provided.



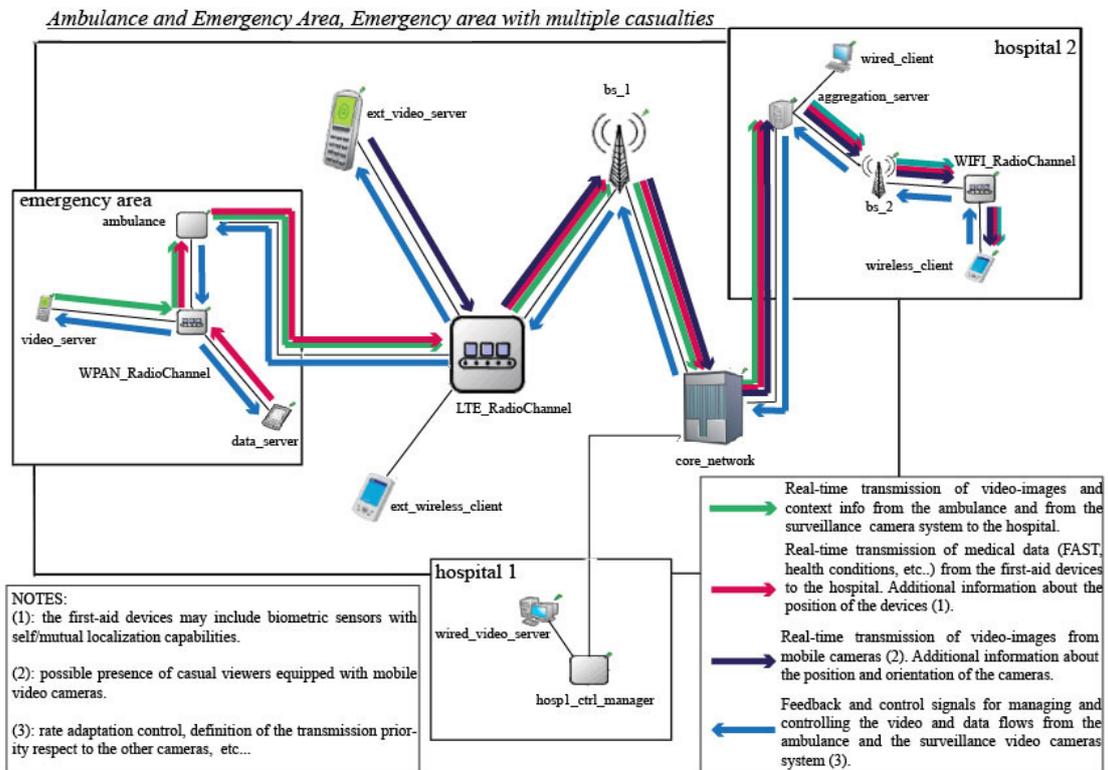


Figure 4.15: Emergency area with multiple casualties: information flows.

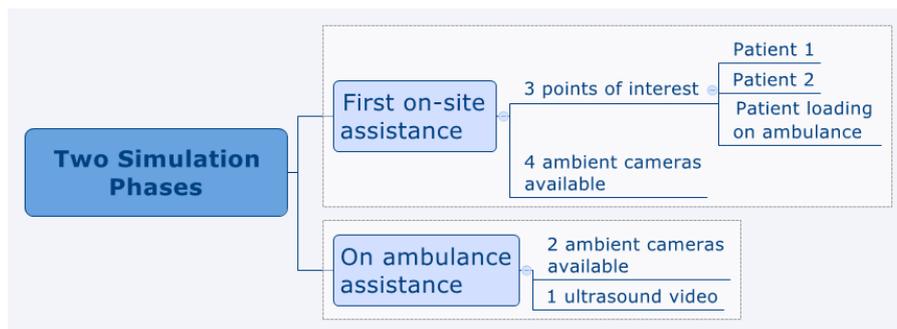


Figure 4.16: Logical phases of UC2 simulations, representing different events and situations within the emergency area.

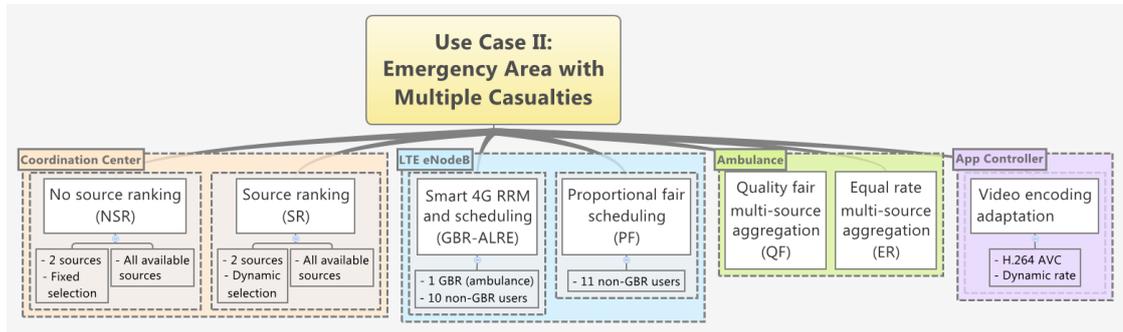


Figure 4.17: Optimization techniques considered for UC2 simulations.

Use Case 2 - Phase 1			
CAMERA	Point of Interest nr.1	Point of Interest nr.2	Point of Interest nr.3q
Ambient Camera 0 ( <i>video_server[0]</i> ) 1024x720, 25fps			
Ambient Camera 1 ( <i>video_server[1]</i> ) 1024x720, 25fps			
Ambient Camera 3 ( <i>video_server[2]</i> ) 1280x960, 25fps			
Ambient Camera 4 ( <i>video_server[3]</i> ) 720x576, 25fps			

Table 4.7: The used video source for simulating UC2 during Phase 1.

Use Case 2 - Phase 2	
Medical Video ( <i>video_server[4]</i> ) 640x480, 25fps	
Ambient Camera 5 ( <i>video_server[5]</i> ) 640x480, 25fps	
Ambient Camera 6 ( <i>video_server[6]</i> ) 640x480, 25fps	

Table 4.8: The used video source for simulating UC2 during Phase 2.

Emergency Area	
Type of radio link	IEEE 802.11g
Channel model	ITU Commercial Area
Modulation	OFDM (QPSK on each subcarrier)
PHY FEC	Punctured convolutional code, $R_c=1/2$
Cyclic Prefix	1/16

Table 4.9: EA parameters for simulating UC2.

Hospital Area	
Type of radio link	IEEE 802.11g
Channel model	ITU Office Area
Modulation	OFDM (QPSK on each subcarrier)
PHY FEC	Punctured convolutional code, $R_c=1/2$
Cyclic Prefix	1/16

Table 4.10: *Hospital\_2* parameters for simulating UC2.

4G Link	
Type of radio link	LTE Uplink
Duplexing Mode	FDD
Modulation	OFDM (adaptive constellation based on AMC)
Multiple Access Type	SC-FDMA
Minimum Allocable Resource Block	1 PRB (=12 subcarriers x 7 OFDM symbols)
PHY FEC	Turbo code (adaptive Rc based on AMC)
Cyclic Prefix	4.69us (short CP)
Adaptive Modulation and Coding (AMC)	Enabled
Radio Resource Management	PF / GBR-ALRE
Ambulance (Tx)	
<i>Channel model</i>	ITU Pedestrian A
<i>Doppler Frequency</i>	1.8Hz
<i>Max TX Power</i>	18dBm
<i>Antenna Gain</i>	2.0dBi
<i>Nr. of Antennas</i>	1
<i>Lognormal Shadowing sigma dB</i>	4dB
<i>Distance to the eNodeB</i>	500m
eNodeB (Rx)	
<i>Antenna Gain</i>	18.0dBi
<i>Nr. of Antennas</i>	2
<i>Cable Loss</i>	2.0dB
<i>Rx Noise Figure</i>	2.0dB
Additional Non Concerto Users (Tx)	
<i>Nr. of Non Concerto Users</i>	10
<i>Channel model</i>	ITU Pedestrian A
<i>Distance to the eNodeB</i>	Between 50m and 1500m

Table 4.11: eNodeB parameters for simulating UC2.

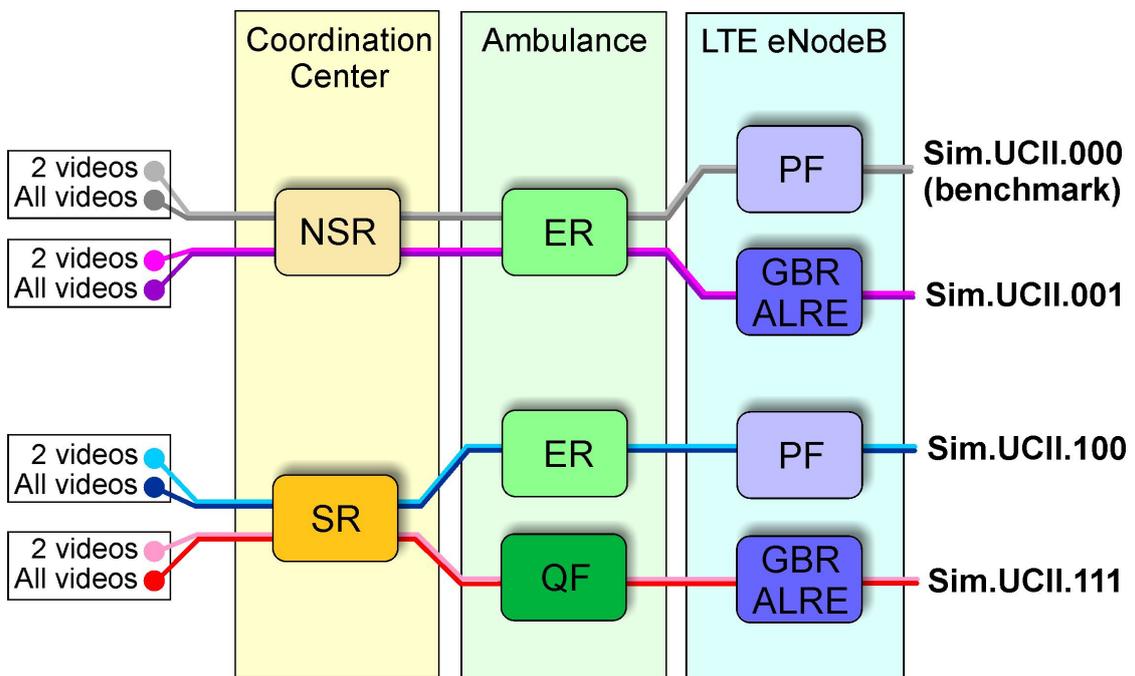


Table 4.12: Scheme of the simulation campaign realized for UC2.

## Case Sim.UCII.000

*Video Encoding Performance: LTE Uplink Performance:*

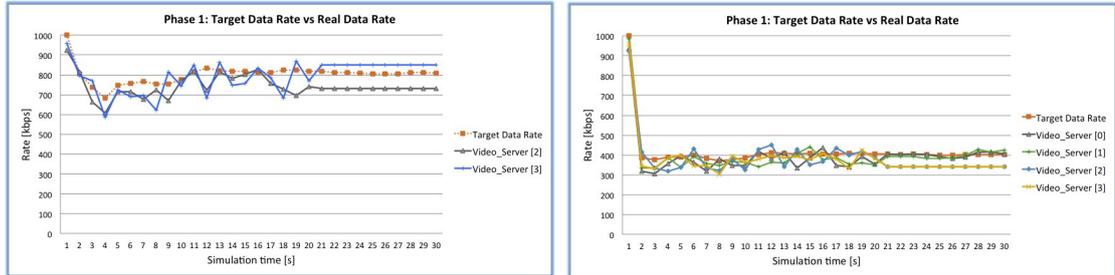


Figure 4.18: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right).

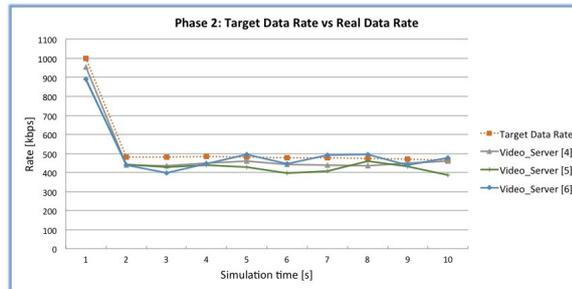


Figure 4.19: Target source rate specified to the APP Controller and achieved rate during Phase 2.

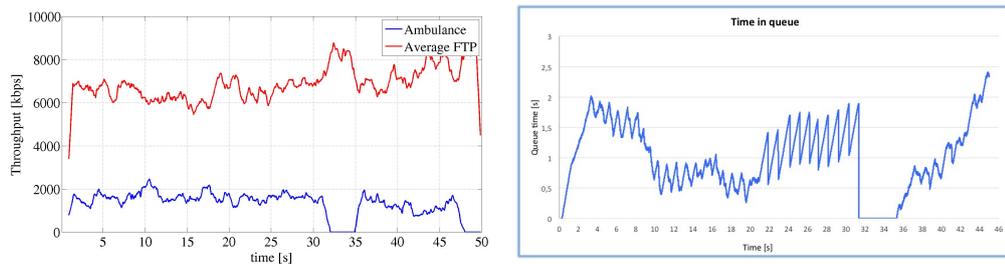


Figure 4.20: Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell (left) and time spent in TX queue at the LTE MAC layer (right). Transmission of 2 videos during Phase 1 (the case with 4 videos provides similar results).

*End-to-end Performance:*

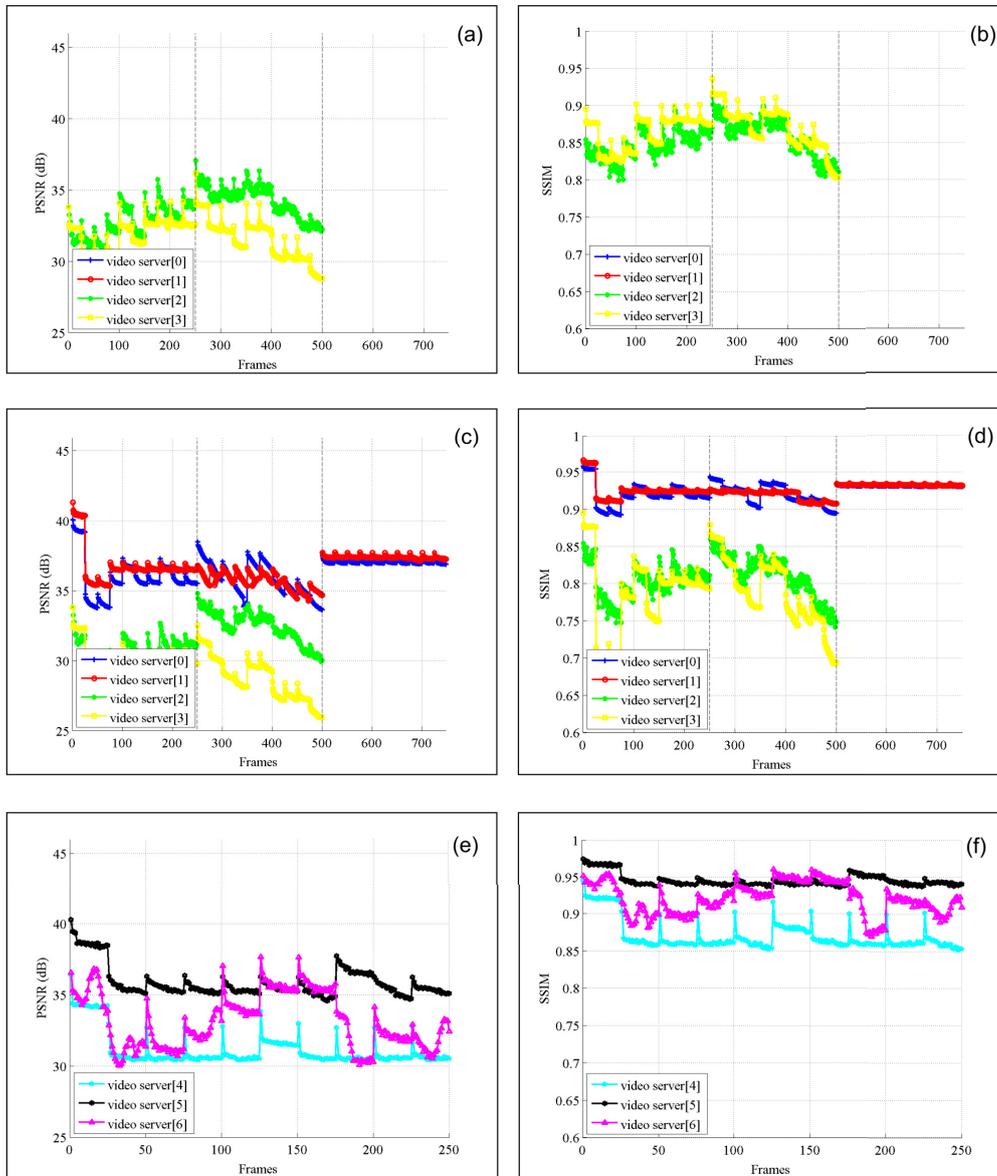


Figure 4.21: Video quality at the doctor’s UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

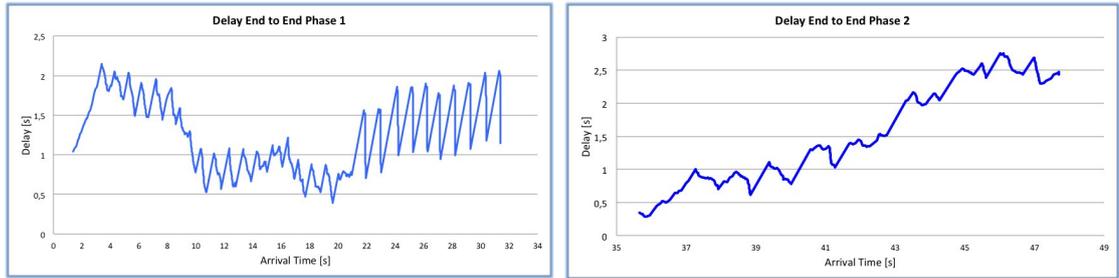


Figure 4.22: End-to-end delay in Phase 1 (left) and Phase 2 (right). The case with 2 videos in Phase 1 has been reported, while similar results have been obtained also in the case with 4 videos.

*Some Remarks:* This case represents the benchmark solution, where no source ranking is performed and equal rate video adaptation is applied within the ambulance. A traditional proportional fair multiuser scheduling is realized at the eNodeB to manage the traffic within the LTE cell. The application controller is enabled, permitting to optimize the video coding based on the available throughput. During Phase 1 the achieved video quality is rather low, especially in case of four transmitted videos, and the quality spread among the sequences is relevant. Finally, the end-to-end delay is quite high and not suitable for a quick interaction with the remote specialist. In Phase 2, the medical video sequence does not reach a quality generally capable to guarantee a satisfactory diagnosis accuracy.

## Case Sim.UCII.001

### Video Encoding Performance:



Figure 4.23: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 3 Mbps at the LTE eNodeB.

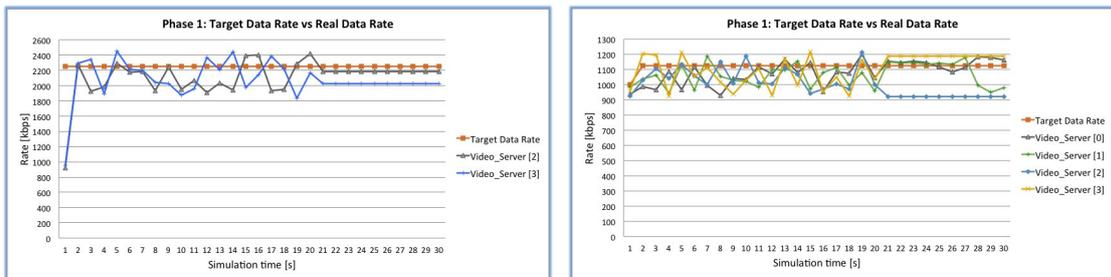


Figure 4.24: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 4.5 Mbps at the LTE eNodeB.



Figure 4.25: Target source rate specified to the APP Controller and achieved rate during Phase 2, with GBR = 3 Mbps (left) and GBR = 4.5 Mbps (right) at the LTE eNodeB.

*LTE Uplink Performance:*

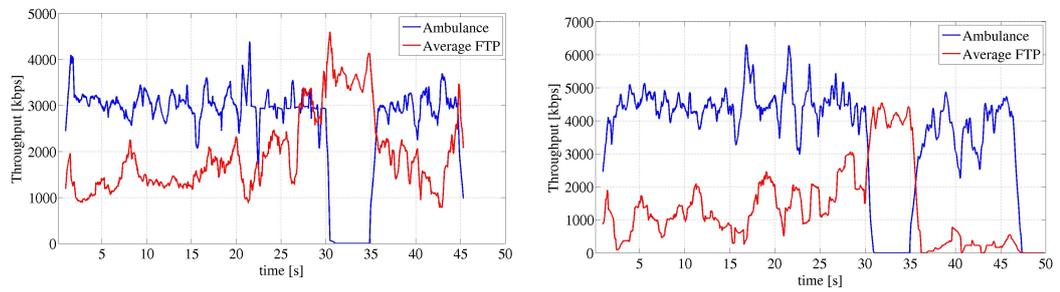


Figure 4.26: Rate allocated by the LTE eNodeB to the ambulance and to additional users in the cell. Transmission of 2 videos during Phase 1, GBR=3 Mbps (left) and 4.5 Mbps (right). Similar results have been obtained in the case with 4 videos in Phase 1.

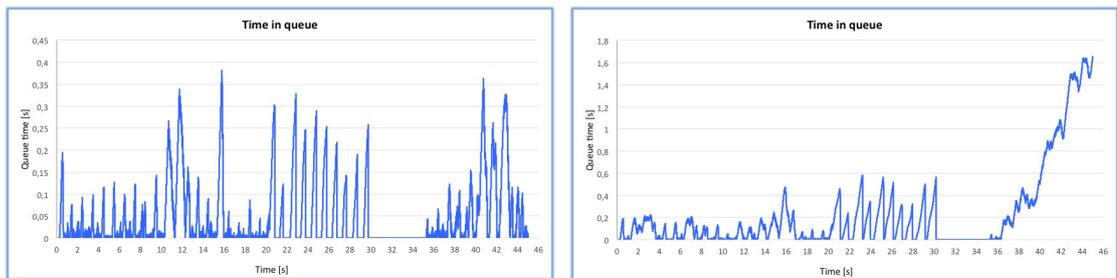


Figure 4.27: Time spent in TX queue at the LTE MAC layer. The case of 2 videos in Phase 1 has been reported, with GBR=3 Mbps (left) and GBR=4.5 Mbps (right) . Similar results have been obtained in the case with 4 videos in Phase 1.

*End-to-end Performance:*

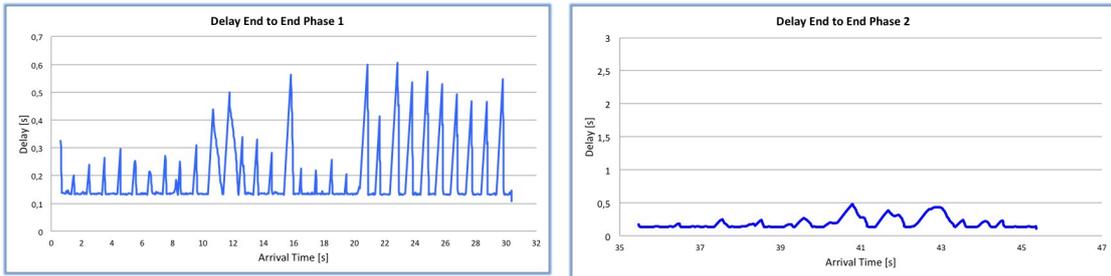


Figure 4.28: End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 3 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1.

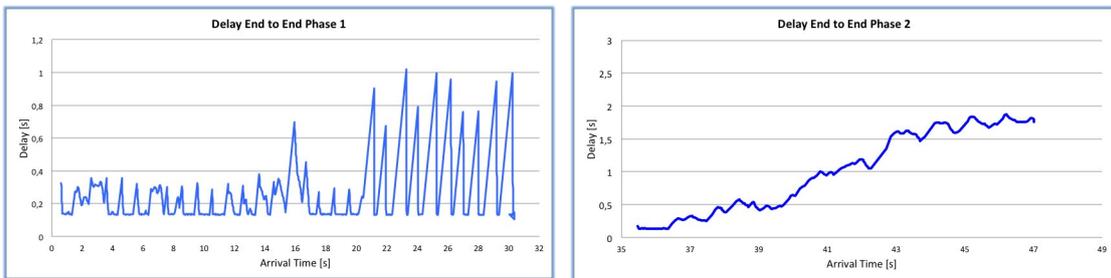


Figure 4.29: End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 4.5 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1.

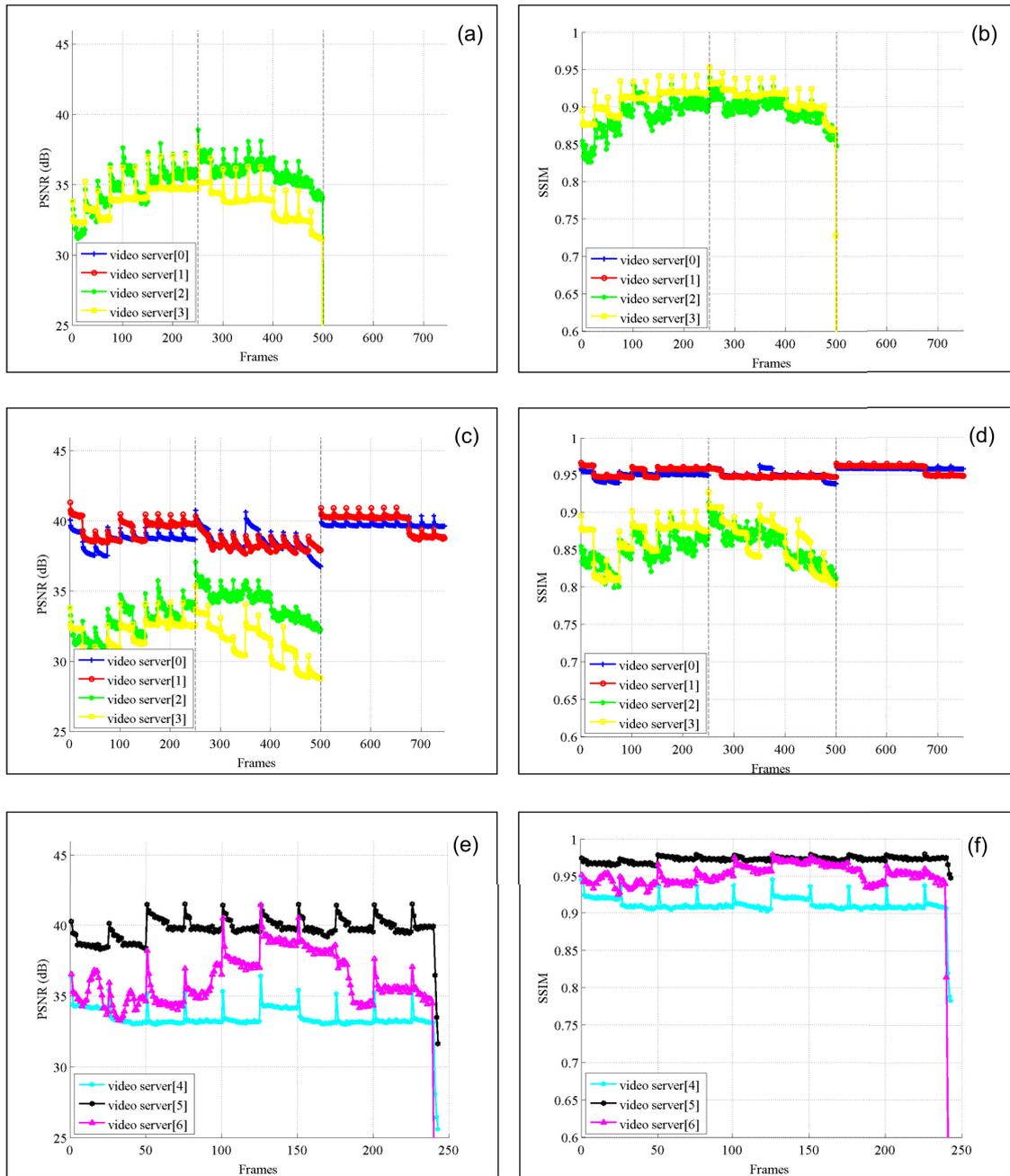


Figure 4.30: GBR = 3 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

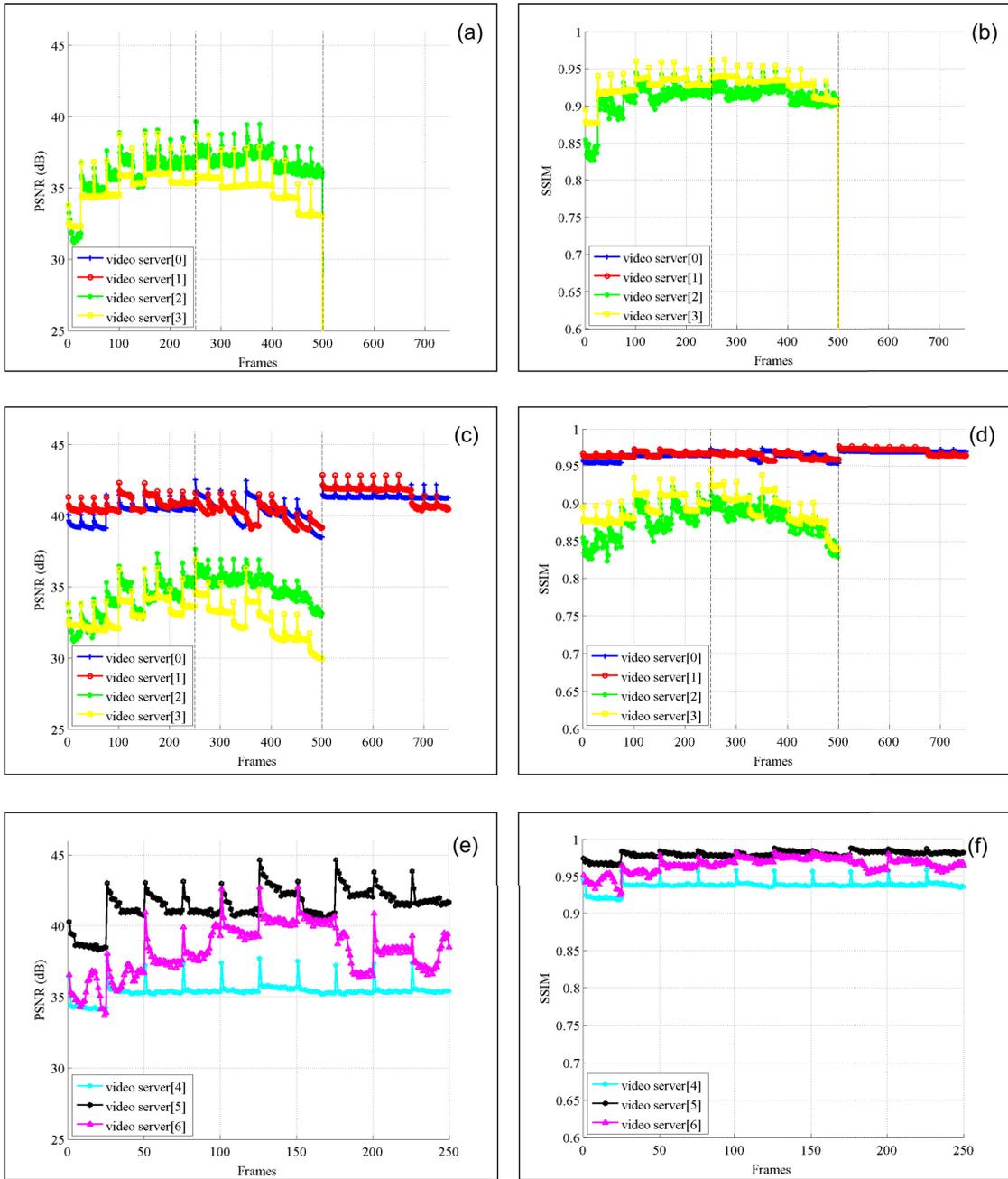


Figure 4.31: GBR = 4.5 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

*Some Remarks:* In the Sim.UCII.001 case, the system intelligence is mainly concentrated in the radio management performed by the LTE eNodeB. As it can be noted from Fig. 4.28 and Fig. 4.29, the delay is considerably lower than with the PF policy and more suitable to real-time or quasi real-time communications. However, also in this case the quality spread among the sequences is considerable, as no quality fairness is guaranteed within the ambulance. In particular, in Phase 2, the ultrasound video has a SSIM still lower than 0.95, considered a target quality level to achieve a good diagnostic service.

## Case Sim.UCII.100

*Video Encoding Performance: LTE Uplink Performance:*

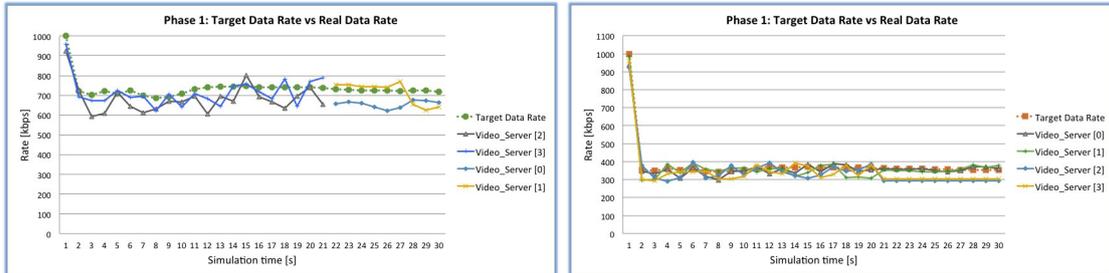


Figure 4.32: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right).

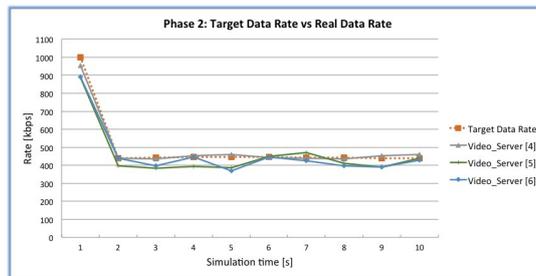


Figure 4.33: Target source rate specified to the APP Controller and achieved rate during Phase 2.

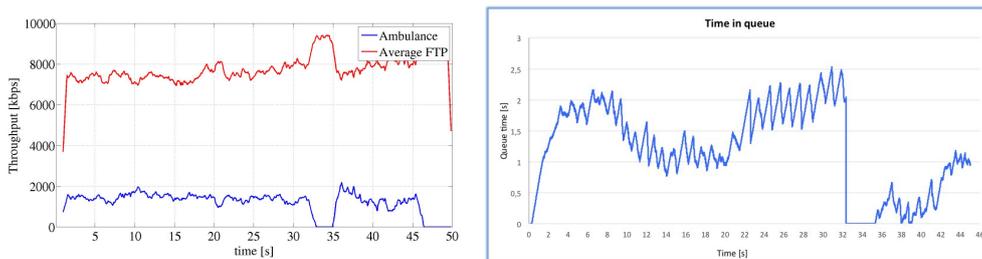


Figure 4.34: Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell (left) and time spent in TX queue at the LTE MAC layer (right). Transmission of 2 videos during Phase 1 (the case with 4 videos provides similar results).

*End-to-end Performance:*

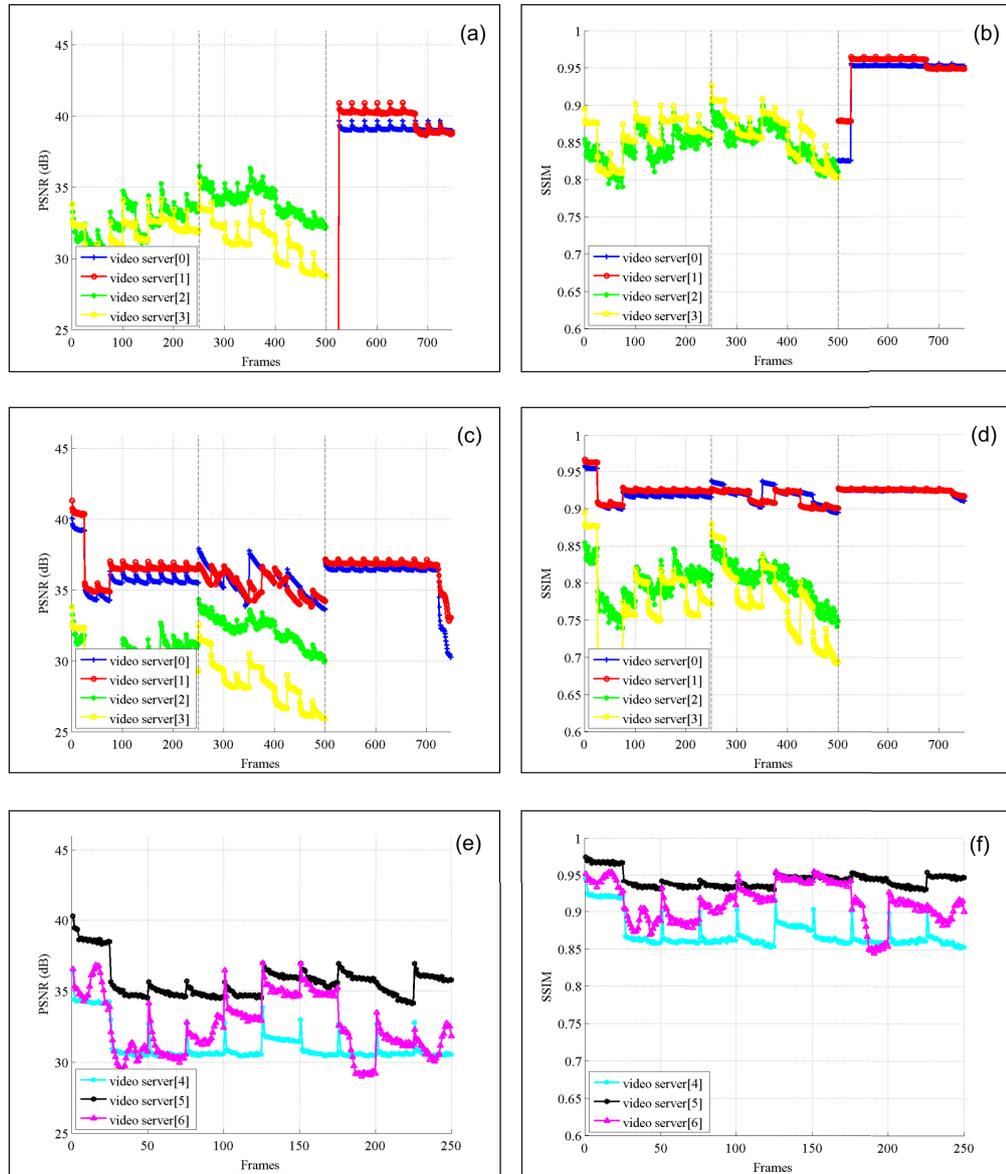


Figure 4.35: Video quality at the doctor’s UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

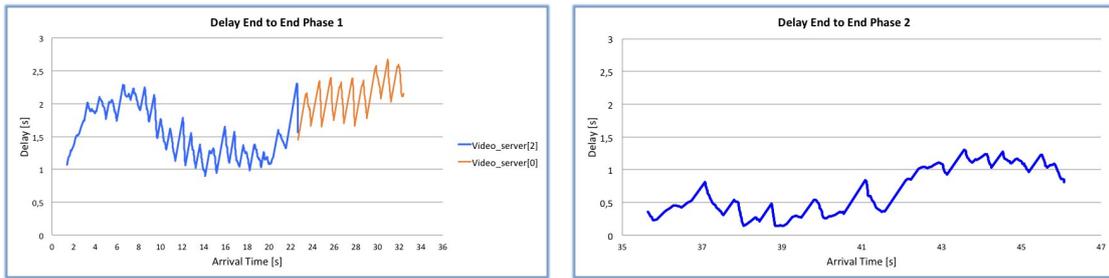


Figure 4.36: End-to-end delay in Phase 1 (left) and Phase 2 (right). The case with 2 videos in Phase 1 has been reported, while similar results have been obtained also in the case with 4 videos.

*Some Remarks:* Unlike the benchmark, in this case source prioritization is applied at the CC and the ranking is communicated to the ambulance. In case of 2 transmitted videos in Phase 1, the transmitted sources are automatically changed after the first 20s, passing from *video\_server[2]* and *video\_server[3]* to *video\_server[0]* and *video\_server[1]*. In fact, in the interval between 20s and 30s (patient loading on the ambulance) the two cameras initially selected are less significant than the other two.

## Case Sim.UCII.111

*Video Encoding Performance:*

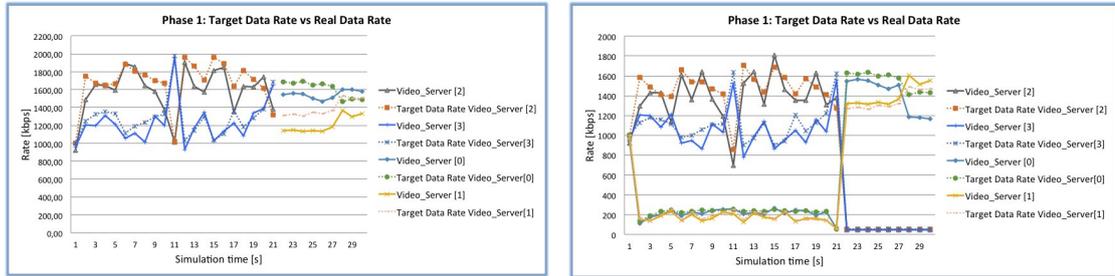


Figure 4.37: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 3 Mbps at the LTE eNodeB.

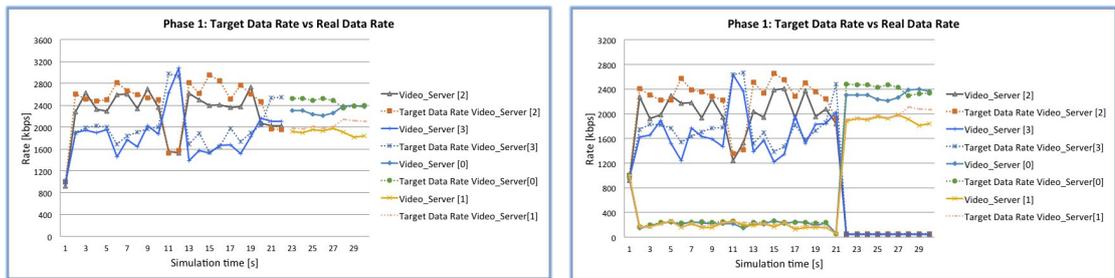


Figure 4.38: Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 4.5 Mbps at the LTE eNodeB.



Figure 4.39: Target source rate specified to the APP Controller and achieved rate during Phase 2, with GBR = 3 Mbps (left) and GBR = 4.5 Mbps (right) at the LTE eNodeB.

*LTE Uplink Performance:*

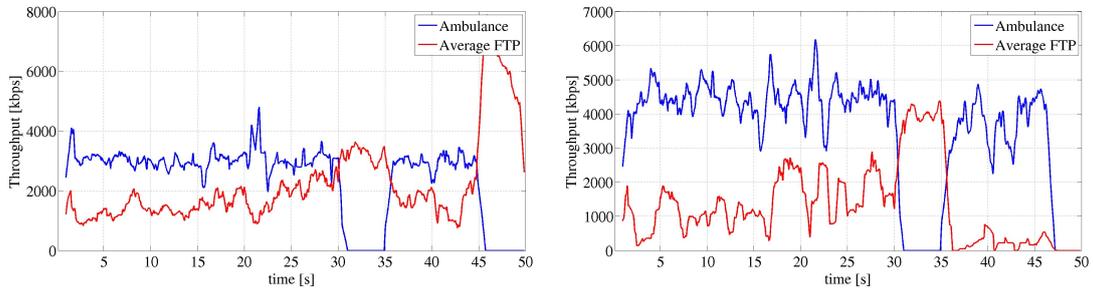


Figure 4.40: . Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell. Transmission of 2 videos in Phase 1, GBR=3 Mbps (left) and 4.5 Mbps (right). Similar results have been obtained in the case with 4 videos in Phase 1.

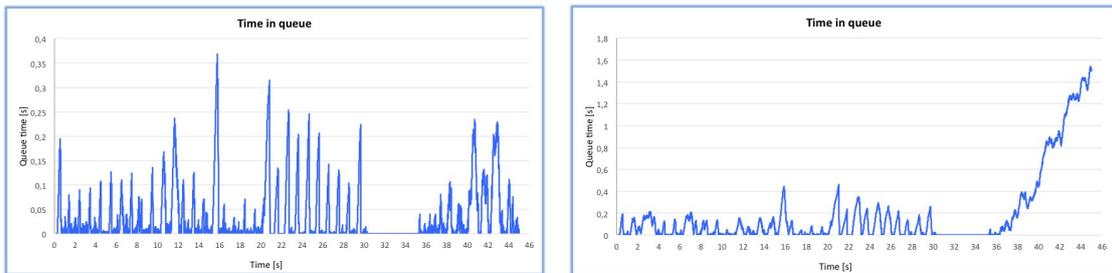


Figure 4.41: Time spent in TX queue at the LTE MAC layer. The case of 2 videos in Phase 1 has been reported, with GBR=3 Mbps (left) and GBR=4.5 Mbps (right) . Similar results have been obtained in the case with 4 videos in Phase 1.

*End-to-end Performance:*

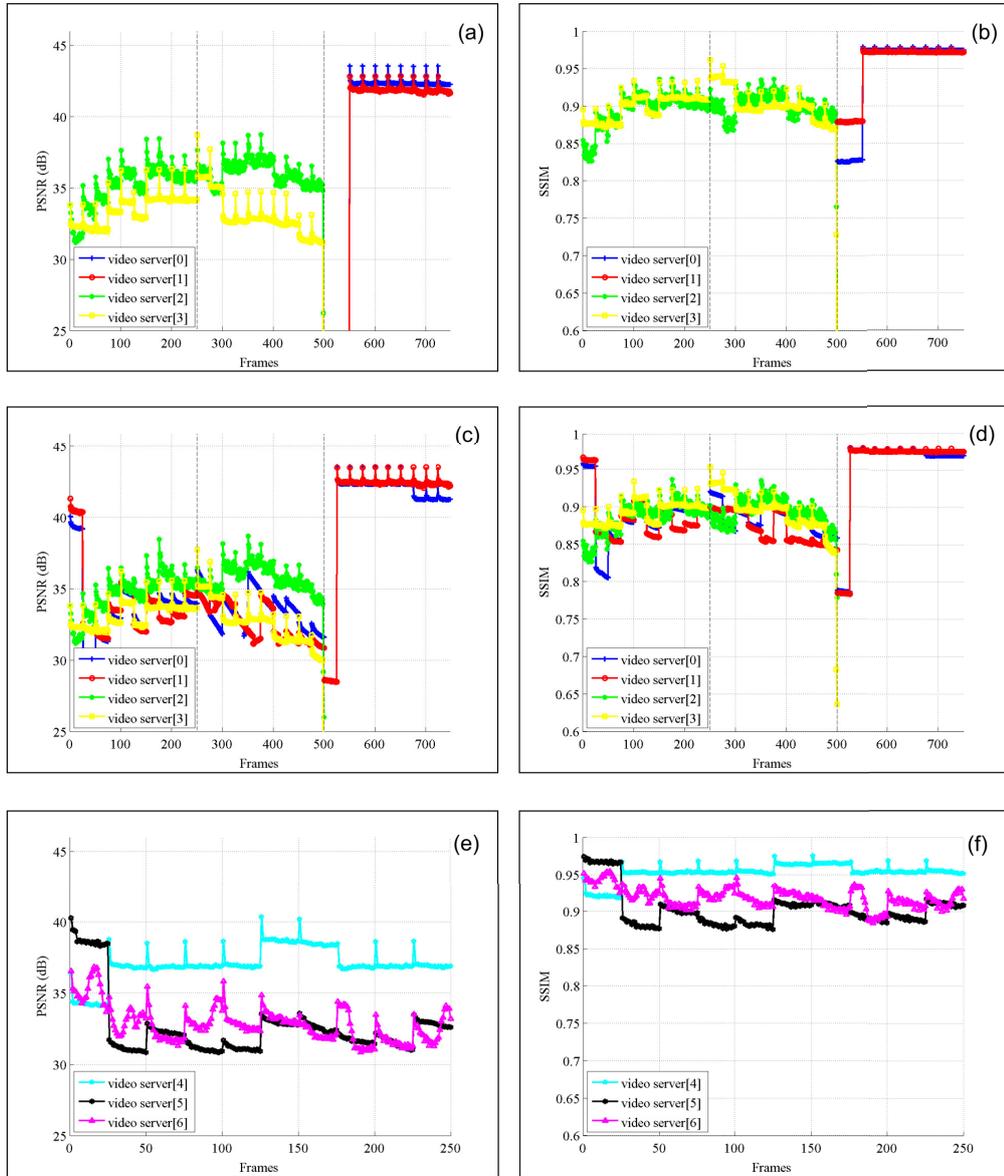


Figure 4.42: GBR = 3 Mbps at the LTE eNodeB. Video quality at the doctor’s UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

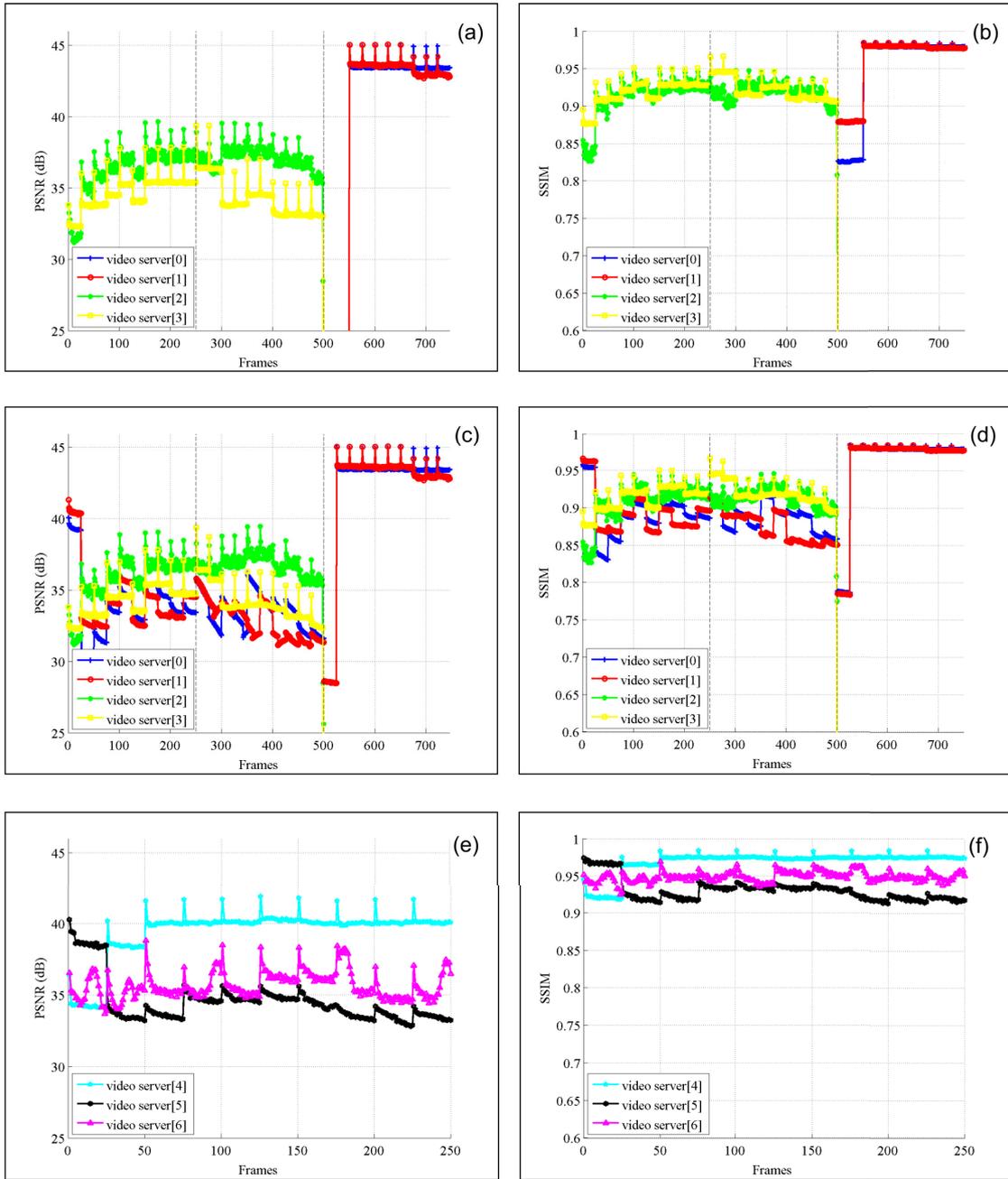


Figure 4.43: GBR = 4.5 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos.

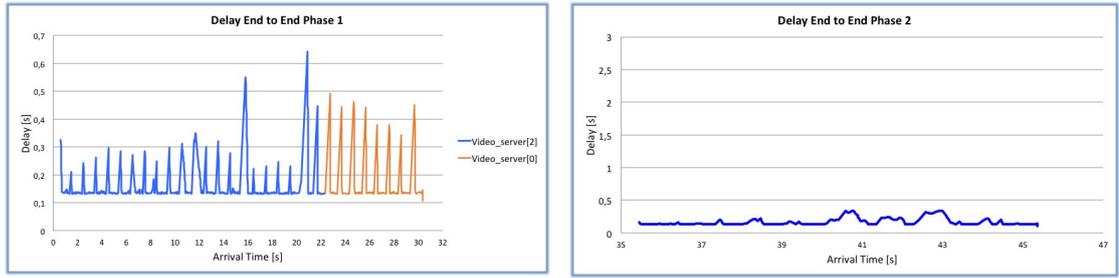


Figure 4.44: End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 3 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1.

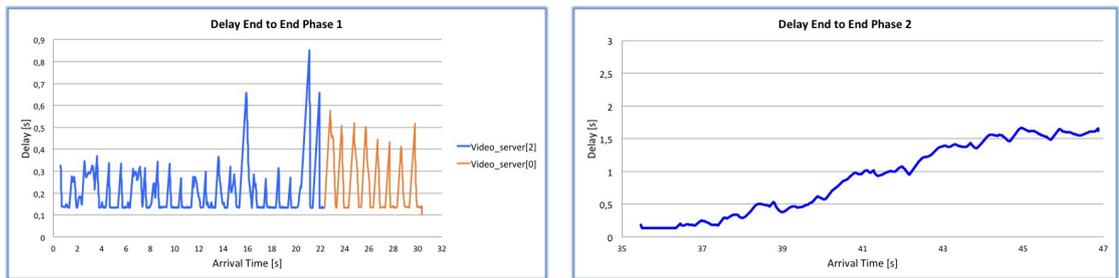


Figure 4.45: End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 4.5 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1.

*Some Remarks:* This case represents the fully optimized solution, including camera ranking at the CC, joint video coding adaptation based on quality fairness criteria and smart radio resource management at the LTE radio access network, capable to guarantee a fixed bit-rate for the CONCERTO user. As it can be noted from the numerical results, the quality spread among the ambient videos is significantly lower than that achieved with a simple equal rate policy. At the same time, the end-to-end delay allows a real-time interaction between the on-site staff and the specialized personnel at the hospital. Finally, it is worth noting that the quality-based video coding adaptation allows to specify a higher quality for the medical video during phase 2, enabling accurate tele-diagnosis service.

## UC2: Numerical Result Comparison

In this paragraph, a comparison of the different schemes simulated for UC2 is reported, in terms of both end-to-end quality and LTE uplink performance. As in previous figures, the File Transfer Protocol (FTP) label corresponds to non-CONCERTO traffic within the cell. In this paragraph, some remarks about the

average performance achieved with the different strategies considered for UC2 are reported, as depicted in Fig. 4.49 to Fig. 4.47.

*End-to-end Performance:* The average PSNR and SSIM achieved with the different transmission schemes is reported in Fig. 4.48 and Fig. 4.49, respectively. First of all, it is remarkable that in all the simulated situations the ambulance is able to estimate its available throughput across the 4G link and to adapt the source rates accordingly, by means of the application controller. Based on the reported comparative results, it is possible to make some observations:

- When a PF scheduling is applied within the eNodeB, the throughput provided to the ambulance is not always sufficient to guarantee good quality. When all videos are transmitted via PF in Phase 1, the corresponding SSIM is distributed between 0.77 and 0.92 (28.68 dB and 36.62 dB, in terms of PSNR). An even higher quality spread is observed when only 2 videos are selected through SR for transmission, due to the low motion of the scene capture by *video\_server[0]* and *video\_server[1]*. More important, with PF the medical ultrasound video (Phase 2) is not able to achieve a video quality sufficient to enable an accurate tele-diagnosis.
- When SR is employed to select and transmit 2 videos, the system is able to automatically track the Point of Interest, switching from *video\_server[2]* and *video\_server[3]* to *video\_server[0]* and *video\_server[1]* during Phase 1. Transmitting only two videos out of the four available in Phase 1 permits to allocate more bandwidth to the most significant sources, improving the corresponding end-to-end quality with respect to the transmission of all the available videos.

The possibility to guarantee a given throughput to the CONCERTO user across the LTE link turns out to be very important to achieve a satisfactory quality, especially when more than 2 videos are contemporary transmitted. For example, if 4.5 Mbps are guaranteed to the ambulance by the eNodeB, all videos are received with average SSIM higher than 0.88 and PSNR not lower than 32.8 dB.

- When the QF mechanism is implemented within the ambulance, the quality spread between the ambient streams is effectively minimized. In particular, the QF rate allocation turns out to be fundamental when a medical stream is multiplexed with the others, in order to achieve the video quality required by the specialists at the hospital. In particular, with all the Sim.111 schemes the ultrasound video (*video\_server[4]*) is received with SSIM higher than 0.95 and PSNR above 37 dB. On the contrary, when the ER policy is applied, the medical video is always received with a quality lower than the ambient videos, due to a lower compression efficiency.

*LTE Uplink Performance:* A comparison of the average throughput achieved in the different schemes of UC2 is given in in Fig. 4.46. In particular, the performance of the ambulance (the CONCERTO UE) and the aggregate throughput of the non-CONCERTO UEs (FTP) is reported. In Fig. 4.47, the 3rd quartile of the packet delay at the LTE link, *i.e.*, the packet delay that is exceeded with probability 25%, is plotted.

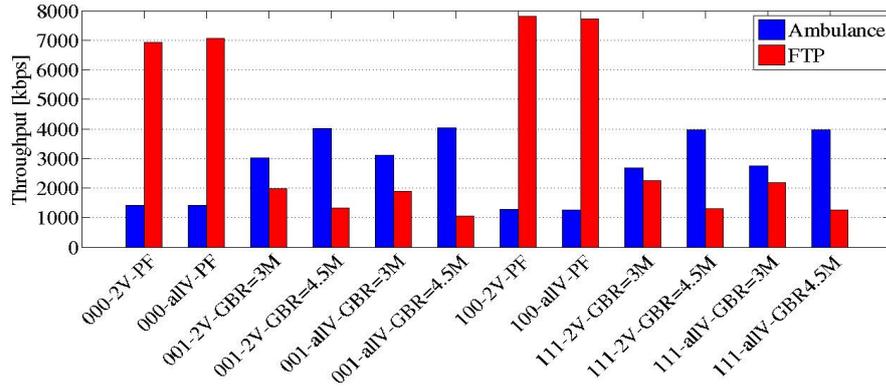


Figure 4.46: UC2: average throughput of the ambulance and FTP users (aggregate) for each simulated scheme.

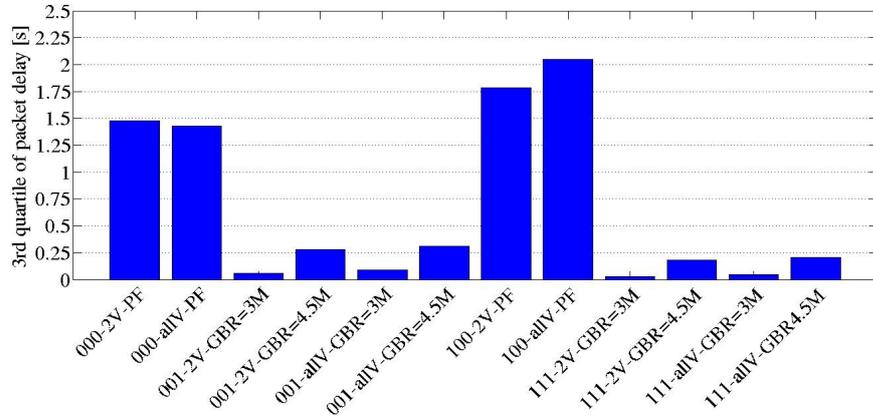


Figure 4.47: UC2: 3<sup>rd</sup> quartile of the packet delay, *i.e.*, 75<sup>th</sup> percentile, of the packet delay CDF for each simulated scheme.

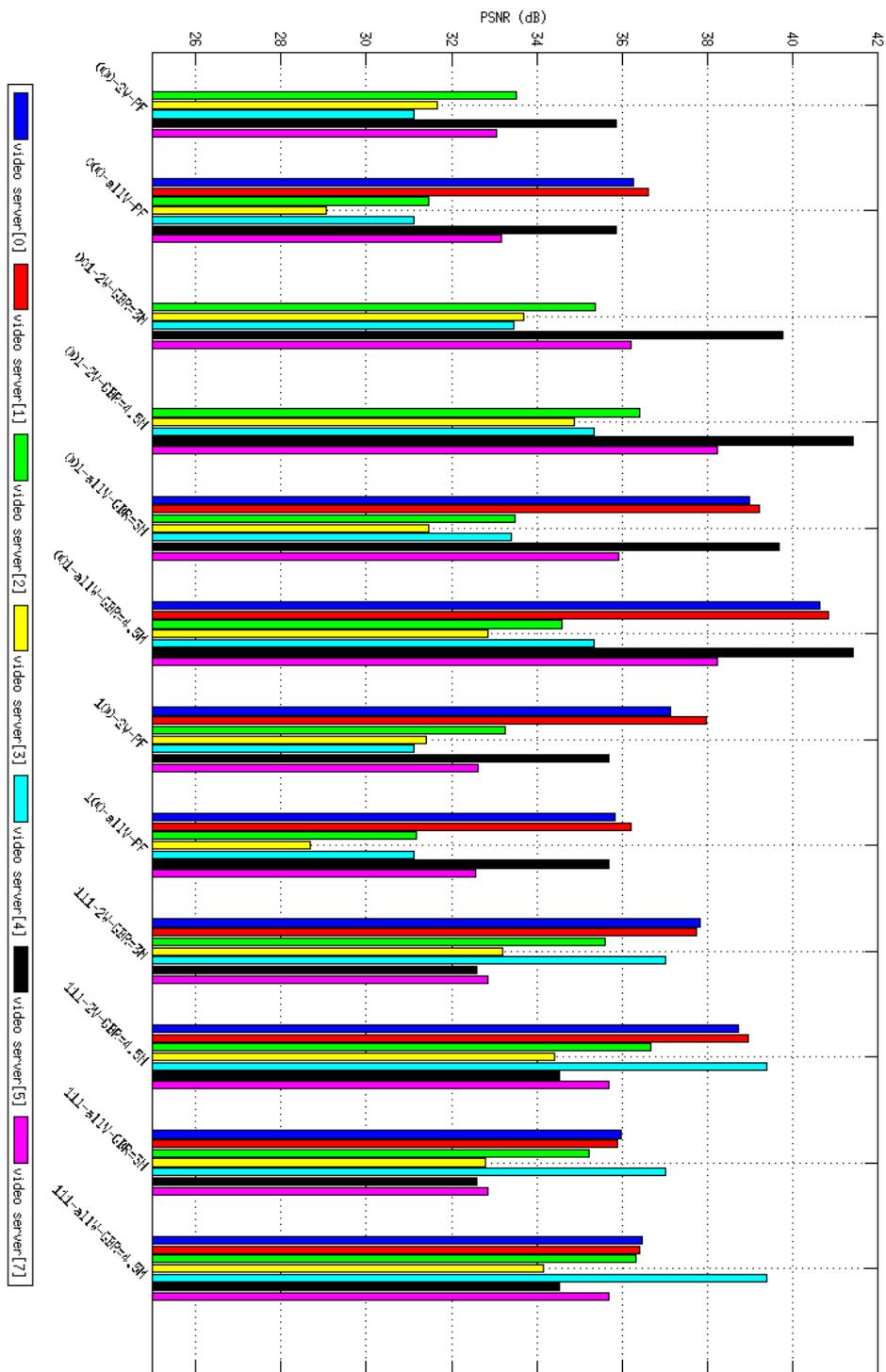


Figure 4.48: UC2: comparison of the average PSNR achieved for the received videos in the different cases.

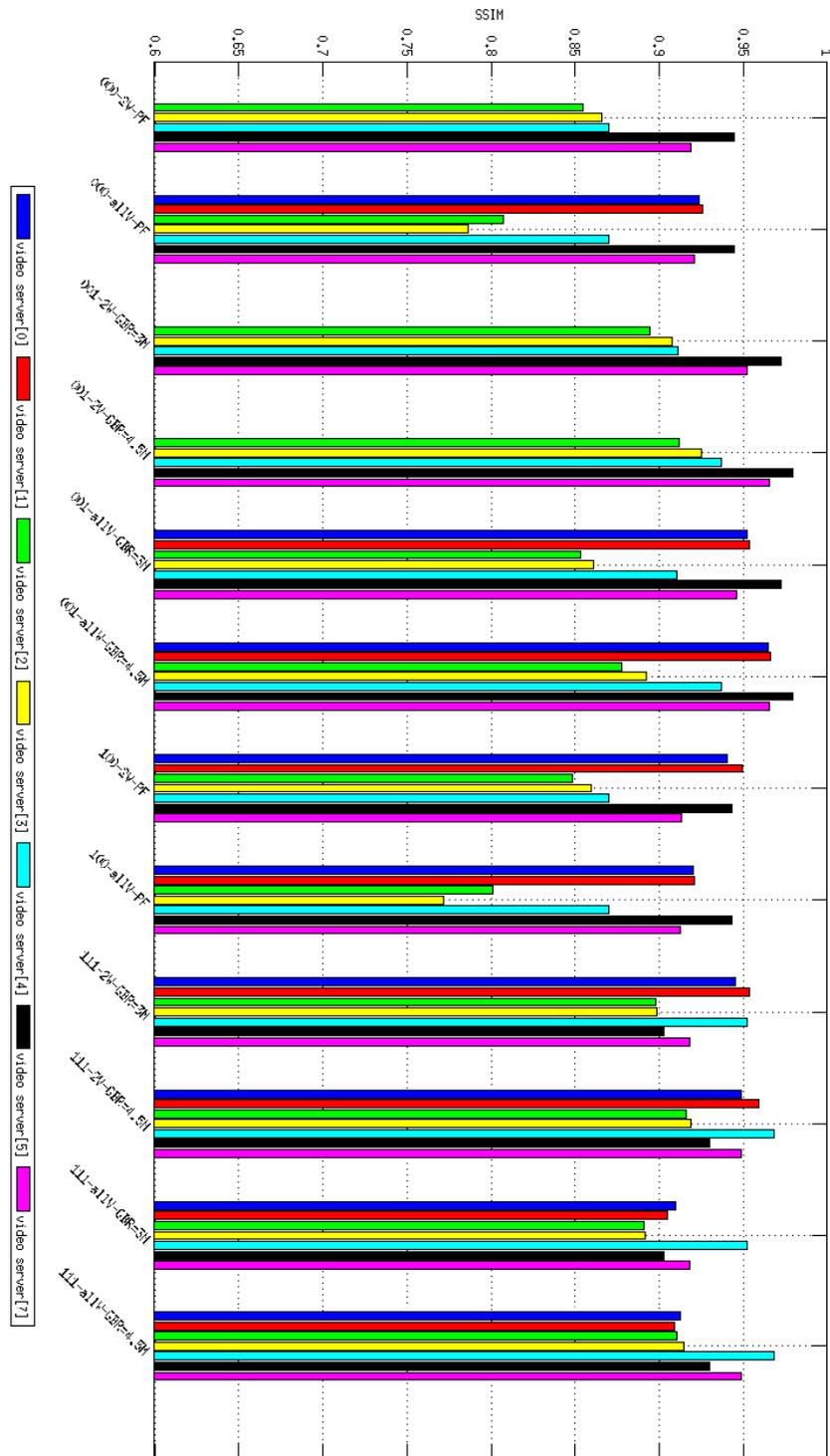


Figure 4.49: UC2: comparison of the average SSIM achieved for the received videos in the different cases.

## 4.5 Conclusions

This Chapter described the preliminary design of the software simulator, an important tool jointly developed by the consortium to evaluate the performance of the proposed algorithms and to validate the overall approach followed in the project CONCERTO. After the description of the general system architecture and the identification of the primary application scenarios on which the simulator is focusing, the main SW modules involved in the communication process have been described in some details, including functional requirements, inputs and outputs. In addition, this Chapter is aimed at validating the proposed solutions through the project simulator developed on the OMNeT++ framework. Several use cases have been addressed showing the benefits of the proposed system in managing different emergency and healthcare situations. Different levels of optimization have been considered, in order to evaluate the potential system performance in case of different extent of CONCERTO in the next future. In particular, it is assumed the presence of a centralized coordination unit (*i.e.*, the CC), capable to prioritize the available ambient and medical streams. Camera ranking techniques are applied at the CC to guarantee the transmission of significant and high quality contents throughout the emergency event. Multi-stream adaptation is performed within the ambulance and source coding optimization is realized by dedicated application controllers. In addition, the benefits deriving from the adoption of advanced multi-user scheduling and resource allocation techniques within the LTE eNodeB, capable to guarantee a fixed bit-rate to the CONCERTO user are evaluated. Different combinations of the above-mentioned solutions have been simulated and compared to a proper benchmark, representing traditional communication systems. The considered use cases have been evaluated addressing different service metrics, *e.g.*, throughput and delay, and video quality indicators, *i.e.*, PSNR and SSIM. Based on these numerical results, the overall benefits of the CONCERTO solutions have been presented in different situations. Several results are also included in [6]

# Conclusions

In this thesis, I designed and validated new techniques for media content fusion, delivery and interactive access, while specifically addressing the strict requirements of healthcare services. One of the most important problem arising in the emergency communication scenarios is related to the huge amount of multimedia information which should be transmitted and managed in conjunction with medical data. The generated streams contain redundant information which could be exploited to increase the probability of a lossless video transmission and, on the other side, this information could be fused and/or selected for delivery in case of limited resources. A proper classification of the collected media information through data fusion techniques may allow remote users to interactively move through the different available cameras acquiring a more precise idea of a given crisis scenario. The principal aim of this dissertation is to provide high QoE for medics, which is a necessary condition for providing flawless medical diagnosing of the highest reliability. To this purpose, a system architecture has been envisaged with the aim at satisfying the requirements of healthcare services and to deploy different multi-user scenarios considered in the real world. Specifically, in Chapter 2 I exploited camera ranking algorithms to logically organize and select multiple multimedia streams in the context of large camera networks. In this regard, new camera ranking strategies for selecting and prioritizing multiple information sources constituted by multiple cameras with different positions and orientations have been conceived. In particular, two novel camera ranking algorithms have been presented. The proposed algorithms have been developed with the aim to find, within a VSN, the camera that best satisfies specific ranking criteria. The proposed algorithms are designed to optimize the QoE for the final user in terms of video quality and significance. To test the camera selection techniques, an indoor multi-camera acquisition system has been arranged and the resulting video sequences has been used to perform some evaluations on the camera ranking algorithms. Moreover, I have presented a further scenario in which multi-camera systems are applied in emergency situations. I considered the transmission of health-related video information from an ambulance to a remote hospital, as a support for tele-consultation and tele-diagnosis services. Numerical results are derived by experimental simulations and

subjective tests which testify the efficiency of the proposed techniques in providing the best final user experience in terms of visual quality in contexts of healthcare and security application scenarios.

The long-distance transmission of health-related information (*e.g.*, from an ambulance to a remote hospital) is a challenging task, due to the variability and the limitations of the mobile radio link. In particular, the transmission of multiple video streams can improve the efficacy of the tele-consultation service, but requires a large bandwidth to meet the desired quality, not always guaranteed by the mobile network. Consequently, the optimization of the wireless network is necessary for transporting the large amounts of both multimedia and medical information, either in the downlink and in the uplink. In Chapter 3, I designed novel adaptation schemes for centralized uplink transmissions. In particular, I propose novel cross-layer adaptation strategies for multiple SVC videos delivered over a single LTE channel, which dynamically adjusts the overall transmitted throughput to meet the actual available bandwidth. The proposed adaptation strategies jointly interact with the camera ranking techniques presented in Chapter 2. Novel MAC level scheduling solutions for the uplink have been investigated, capable to differently prioritise the data traffic in a shared wireless network and enabling interactive healthcare applications. The proposed solutions are capable to guarantee a good end-to-end video quality despite fluctuations in the available rate and heavy traffic conditions within the LTE cell. Furthermore, several numerical results are presented showing the end-to-end video quality performance, for both diagnostic and multiple ambient videos. In particular, two novel content-aware aggregation and adaptation strategies and the joint application of the camera ranking algorithms (Chapter 2) are presented. The proposed schemes are suitable for the transmission of a set of SVC videos transmitted over LTE uplink in a mHealth emergency scenario while the context-aware camera selection algorithms can select one or more cameras taking into account specific ranking criteria mainly related to the quality of the visual representation of the object of interest. The proposed adaptation algorithm is based on modeling and evaluation of SSIM quality metric and has been tested in two operational steps. Several numerical results are presented showing the end-to-end video quality performance, for both diagnostic and ambient videos, which results to be optimal for both diagnosis and coordination purposes.

To implement, validate and demonstrate the proposed functionalities, Chapter 4 has been dedicated on the integration of the presented solutions on a common system simulator which is been developed within the CONCERTO project. In particular, I focused my work on the design of the simulation architecture, as well as on the description of the main modules that have been included in the software framework to implement the camera ranking algorithms, the SVC adap-

tation techniques and the multi-user resource allocation and scheduling schemes for LTE uplinks proposed in Chapter 3. Moreover, in Chapter 4 I described the preliminary design of the software simulator, an important tool jointly developed by the consortium to evaluate the performance of the proposed algorithms and to validate the overall approach followed in the CONCERTO project. After the description of the general system architecture and the identification of the primary application scenarios on which the simulator is focusing, the main SW modules involved in the communication process have been described in some details, including functional requirements, inputs and outputs. The simulator structure has been presented, showing the relationship with the architecture and detailing the implementation of the most significant modules. Input and output information, as well as a technical reference to the documents describing the implemented algorithms have been included. Several use cases have been addressed showing the benefits of the proposed system in managing different emergency and healthcare situations. Different levels of optimization have been considered, in order to evaluate the potential system performance in case of different extent in the next future. In particular, I have assumed the presence of a centralized coordination unit (*i.e.*, the CC), capable to prioritize the available ambient and medical streams. Camera ranking techniques are applied at the CC to guarantee the transmission of significant and high quality contents throughout the emergency event. Multi-stream adaptation is performed within the ambulance and source coding optimization is realized by dedicated application controllers. In addition, I have evaluated the benefits deriving from the adoption of advanced multi-user scheduling and resource allocation techniques within the LTE eNodeB, capable to guarantee a fixed bit-rate to the CONCERTO user. Different combinations of the above-mentioned solutions have been simulated and compared to a proper benchmark, representing traditional communication systems. The considered use cases have been evaluated addressing different service metrics, *e.g.*, throughput and delay, and video quality indicators, *i.e.*, PSNR and SSIM. Based on these numerical results, the overall benefits of the presented solutions have been presented in different situations.



# List of Tables

1.1	UC1 summary. . . . .	13
1.2	Use case 2 summary. . . . .	19
2.1	The ranking criteria taken into account by the two proposed algorithms. . . . .	26
2.2	Set of cameras used during the test phase. In evidence the measured FoVs and the resolution. . . . .	32
2.3	Set of cameras used for testing the CRDS. . . . .	37
2.4	The $C_1$ parameters and the linear pixel per meter $q_{lin}$ at a distance $d' = 5$ m. . . . .	38
2.5	PMOS results. . . . .	40
2.6	Subjective test results (first column) and camera ranking results for the implemented techniques. . . . .	42
2.7	Distances and correlation between automatic and human ranking with $\bar{v} = 5.7$ m/s. . . . .	44
2.8	Some parameters of the camera involved in the emergency scenario. . . . .	47
2.9	The camera ranking algorithms results. . . . .	47
3.1	M-health user GBR $\bar{R}_0$ , average rate of e-health user $R_0$ and average rate of the best-effort users $R_k$ , $k \in \mathcal{K}_2$ . . . . .	69
3.2	List of most used symbols and acronyms. . . . .	84
3.3	Simulation parameters. . . . .	93
3.4	Position and resolution of the acquisition devices. . . . .	94
3.5	CR results ( $q_v$ ) for the three considered positions of the objects of interest. . . . .	95
4.1	The used video source for simulating the UC1. . . . .	122
4.2	EA parameters for simulating the UC1. . . . .	122
4.3	<i>Hospital_2</i> parameters for simulating the UC1. . . . .	123
4.4	eNodeB parameters for simulating the UC1. . . . .	124
4.5	Set of Simulations of UC1. . . . .	124

4.6	The cameras listed in red are selected for transmission from the emergency area when the source ranking is activated in Phase 1. . . . .	126
4.7	The used video source for simulating UC2 during Phase 1. . . . .	133
4.8	The used video source for simulating UC2 during Phase 2. . . . .	134
4.9	EA parameters for simulating UC2. . . . .	134
4.10	<i>Hospital_2</i> parameters for simulating UC2. . . . .	134
4.11	eNodeB parameters for simulating UC2. . . . .	135
4.12	Scheme of the simulation campaign realized for UC2. . . . .	136

# List of Figures

1.1	Ambulance and emergency area use case. . . . .	9
1.2	Emergency area with multiple casualties use case. . . . .	14
2.1	Example of VSN. . . . .	27
2.2	The pinhole camera model: A) 3D representation; B) 2D representation. . . . .	28
2.3	(a) Camera network composed by two nodes. (b) 3D representation of the camera network designed for the test session. . . . .	32
2.4	Frames acquired by the selected cameras. In evidence the differentiation of the target position on which the camera selection is computed. . . . .	33
2.5	The camera selection result for the different target positions. . . . .	35
2.6	Video frames acquired by the selected cameras based on the target position. For each camera, the initial and the final captured target position is highlighted. . . . .	36
2.7	Left: Frame extracted from $C_1$ . Right: the related AoI. . . . .	37
2.8	Distances (KTD, C-SCC, and WKTD) between CRDS and the subjective tests, as a function of the parameter $\bar{v}$ in (2.9), for the recorded scene. . . . .	43
2.9	The multi camera simulator main interface. . . . .	44
2.10	An example of camera ranking result. . . . .	45
2.11	Left: the reproduction of the emergency scenario. Right: the simulated scenario version. . . . .	46
2.12	The scene of interest captured by the different cameras. . . . .	48
2.13	The selected cameras by the proposed algorithms. . . . .	49
2.14	The camera ranking algorithms results for the whole emergency area. . . . .	50
2.15	The re-sampling technique suitable for a fair comparison between different Scene of Interest cameras. . . . .	51
2.16	The different implemented techniques for the spatial video up/down-sampling. . . . .	52
3.1	The proposed mHealth architecture for emergency scenarios. . . . .	62

3.2	Average PSNR received at the hospital when a GBR $\bar{R}_k = 4$ Mbps is guaranteed to the Ambulance. . . . .	67
3.3	Average PSNR received at the hospital when different GBR are guaranteed to the Ambulance. . . . .	69
3.4	The proposed mHealth architecture for emergency scenarios. . . . .	71
3.5	The camera ranking algorithm results for the whole emergency area. . . . .	73
3.6	Received SSIM averaged over each adaptation interval, when transmitting the three outdoor video sequences in the first stage, for different values of GBR granted to the ambulance. The quality weights are set to 1 for all videos. . . . .	77
3.7	SSIM averaged over each adaptation interval received at the hospital, when transmitting all the three (a) and two (b) outdoor video sequences in the first stage. GBR $R_0 = 1.0$ Mbps. The quality weights are set to 1 for all videos. . . . .	78
3.8	SSIM averaged over each adaptation interval received at the hospital, when transmitting the two indoor ambient video sequences and the ultrasonography video in the second stage, for different values of GBR granted to the ambulance. The quality weights are set to 1 for the ambient videos, and 2 for the ultrasonography video. . . . .	79
3.9	The proposed m-health architecture for emergency scenarios. . . . .	80
3.10	Block diagram reporting the exchange of information in the proposed multi-camera management system. . . . .	82
3.11	(a) the 2D pinhole camera model. In evidence the FoV, the camera orientation and the position of an object of interest. (b) the camera ranking algorithm results for the whole emergency area. . . . .	83
3.12	Utility models in terms of the SSIM. . . . .	86
3.13	Available cameras in phase 1 ( <i>Cameras 1-5</i> ) and phase 2 ( <i>Cameras 6-7</i> and <i>Ultrasound</i> ). . . . .	96
3.14	3D plots of the normalized weighted sum of the utilities $\sum_{v \in \mathcal{V}_x} w_v U_v(F_v)$ , and the normalized sum of the utilities $\frac{1}{x} \sum_{v \in \mathcal{V}_x} U_v(F_v)$ with respect to the parameter $\alpha$ and the number of transmitted video $x$ . The utilities are obtained from the inner maximization of problem (3.15) with $\bar{R}_0 = 3$ Mbps. The markers in (a) indicate the optimal solutions for $x$ . . . . .	97
3.15	3D plots of the optimum number of transmitted videos $x^*$ , and the normalized sum of the utilities $\frac{1}{x^*} \sum_{v \in \mathcal{V}_{x^*}} U_v(F_v)$ with respect to the parameter $\alpha$ and the GBR granted to the Ambulance. . . . .	97

3.16	Expected and received SSIM at destination (hospital) averaged over each adaptation interval for the five outdoor video sequences transmitted in the first phase of the emergency scenario, when a GBR equal to 3 Mbps is granted to the ambulance. . . . .	98
3.17	SSIM at destination (hospital) averaged over each adaptation interval for the ultrasound and the two ambient video sequences of the ambulance transmitted in the second phase of the emergency scenario, for different values of GBR granted to the ambulance. The quality weights are set to 2 for the Ultrasound video and to 1 for the ambient videos. . . . .	100
4.1	A screenshot of the OMNeT++ integrated environment adopted to develop the CONCERTO simulator. . . . .	106
4.2	General simulation architecture. . . . .	107
4.3	The structure of <i>hospital_2</i> macro module. . . . .	108
4.4	The structure of the <i>Em_area</i> macro module. . . . .	110
4.5	General architecture of the video sources ( <i>e.g.</i> , mobile phone, smart camera, etc.), single- and multi-access client terminals ( <i>e.g.</i> , remote doctor's PC, tablet, etc.). . . . .	111
4.6	Structure of the phy layer module. . . . .	112
4.7	Ambulance and emergency area: information flows. . . . .	120
4.8	End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	125
4.9	End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	126
4.10	End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	127
4.11	End-to-end PSNR for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	128
4.12	End-to-end SSIM for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	129
4.13	End-to-end SSIM for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	130
4.14	End-to-end SSIM for RTP-based simulations. Comparison of results obtained with and without PL-FEC protection. . . . .	131
4.15	Emergency area with multiple casualties: information flows. . . . .	132
4.16	Logical phases of UC2 simulations, representing different events and situations within the emergency area. . . . .	132
4.17	Optimization techniques considered for UC2 simulations. . . . .	133
4.18	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right). . .	137

4.19	Target source rate specified to the APP Controller and achieved rate during Phase 2. . . . .	137
4.20	Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell (left) and time spent in TX queue at the LTE MAC layer (right). Transmission of 2 videos during Phase 1 (the case with 4 videos provides similar results). . . . .	137
4.21	Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . .	138
4.22	End-to-end delay in Phase 1 (left) and Phase 2 (right). The case with 2 videos in Phase 1 has been reported, while similar results have been obtained also in the case with 4 videos. . . . .	139
4.23	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 3 Mbps at the LTE eNodeB. . . . .	140
4.24	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 4.5 Mbps at the LTE eNodeB. . . . .	140
4.25	Target source rate specified to the APP Controller and achieved rate during Phase 2, with GBR = 3 Mbps (left) and GBR = 4.5 Mbps (right) at the LTE eNodeB. . . . .	140
4.26	Rate allocated by the LTE eNodeB to the ambulance and to additional users in the cell. Transmission of 2 videos during Phase 1, GBR=3 Mbps (left) and 4.5 Mbps (right). Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	141
4.27	Time spent in TX queue at the LTE MAC layer. The case of 2 videos in Phase 1 has been reported, with GBR=3 Mbps (left) and GBR=4.5 Mbps (right) . Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	141
4.28	End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 3 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	142
4.29	End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 4.5 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	142

4.30	GBR = 3 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . . .	143
4.31	GBR = 4.5 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . . .	144
4.32	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right). . . . .	146
4.33	Target source rate specified to the APP Controller and achieved rate during Phase 2. . . . .	146
4.34	Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell (left) and time spent in TX queue at the LTE MAC layer (right). Transmission of 2 videos during Phase 1 (the case with 4 videos provides similar results). . . . .	146
4.35	Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . . .	147
4.36	End-to-end delay in Phase 1 (left) and Phase 2 (right). The case with 2 videos in Phase 1 has been reported, while similar results have been obtained also in the case with 4 videos. . . . .	148
4.37	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 3 Mbps at the LTE eNodeB. . . . .	149
4.38	Target source rate specified to the APP Controller and achieved rate during Phase 1. Case of 2 videos (left) and 4 videos (right), with GBR = 4.5 Mbps at the LTE eNodeB. . . . .	149
4.39	Target source rate specified to the APP Controller and achieved rate during Phase 2, with GBR = 3 Mbps (left) and GBR = 4.5 Mbps (right) at the LTE eNodeB. . . . .	149

4.40	. Rate allocated by the LTE eNodeB to the ambulance and to the additional users in the cell. Transmission of 2 videos in Phase 1, GBR=3 Mbps (left) and 4.5 Mbps (right). Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	150
4.41	Time spent in TX queue at the LTE MAC layer. The case of 2 videos in Phase 1 has been reported, with GBR=3 Mbps (left) and GBR=4.5 Mbps (right) . Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	150
4.42	GBR = 3 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . . .	151
4.43	GBR = 4.5 Mbps at the LTE eNodeB. Video quality at the doctor's UE within the remote hospital. PSNR values are reported on the left, while SSIM are on the right. Plots (a)-(d) refer to Phase 1, while plots (e) and (f) to Phase 2. The curves reported in (a) and (b) correspond to the case of 2 videos in Phase 1, while (c) and (d) to the case of 4 transmitted videos. . . . .	152
4.44	End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 3 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	153
4.45	End-to-end delay in Phase 1 (left) and Phase 2 (right). The 2 video case has been reported, with a GBR = 4.5 Mbps at the LTE radio link. Similar results have been obtained in the case with 4 videos in Phase 1. . . . .	153
4.46	UC2: average throughput of the ambulance and FTP users (aggregate) for each simulated scheme. . . . .	155
4.47	UC2: 3 <sup>rd</sup> quartile of the packet delay, <i>i.e.</i> , 75 <sup>th</sup> percentile, of the packet delay CDF for each simulated scheme. . . . .	155
4.48	UC2: comparison of the average PSNR achieved for the received videos in the different cases. . . . .	156
4.49	UC2: comparison of the average SSIM achieved for the received videos in the different cases. . . . .	157

# Bibliography

- [1] S. Moretti, S. Cicalò, M. Mazzotti, V. Tralli, and M. Chiani, “Content/context-aware multiple camera selection and video adaptation for the support of m-health services,” *Procedia Computer Science*, vol. 40, pp. 206–213, 2014.
- [2] S. Moretti, M. Mazzotti, and M. Chiani, “Video selection for visual sensor networks: a motion-based ranking algorithm,” 2016, submitted to: IEEE 24st European Signal Processing Conference (EUSIPCO).
- [3] S. Cicalò, M. Mazzotti, S. Moretti, V. Tralli, and M. Chiani, “Cross-layer optimization for m-health svc multiple video transmission over LTE uplink,” in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*. IEEE, 2013, pp. 212–217.
- [4] S. Cicalò, M. Mazzotti, S. Moretti, V. Tralli, and M. Chiani, “An optimization framework for content/context-aware svc multiple video delivery in m-health emergency applications,” 2016, submitted to: IEEE Transaction on Multimedia, Special Issue on Multimedia-based Healthcare.
- [5] M. Mazzotti, S. Moretti, and M. Chiani, “Multiuser resource allocation with adaptive modulation and ldpc coding for heterogeneous traffic in ofdma downlink,” *Communications, IEEE Transactions on*, vol. 60, no. 10, pp. 2915–2925, 2012.
- [6] L. Iacobelli, G. Panza, E. Piri, J. Vehkaperä, M. Mazzotti, S. Moretti, S. Cicalò, L. Bokor, N. Varga, and M. Martini, “An architecture for m-health services: the concerto project solution,” in *Networks and Communications (EuCNC), 2015 European Conference on*. IEEE, 2015, pp. 118–122.

# CONCERTO Deliverables

- [7] “CONCERTO deliverable 2.1: Use cases, usage models and related requirements.”
- [8] “CONCERTO deliverable 2.2: Specification of cross layer system architecture.”
- [9] “CONCERTO deliverable 4.1: Description and evaluation of context-awareness techniques.”
- [10] “CONCERTO deliverable 4.2: Intermediate evaluation and specifications of content adaptation solutions.”
- [11] “CONCERTO deliverable 4.3: Final evaluation of context-awareness solutions.”
- [12] “CONCERTO deliverable 6.3: Concerto simulator: Final architecture and numerical result.”
- [13] “CONCERTO deliverable 5.1: Comparative study of cooperative communications and networking solutions for mshtm applications.”
- [14] “CONCERTO deliverable 6.1: Preliminary design and implementation report of the system simulator.”

# References

- [15] L. Skorin-Kapov and M. Matijasevic, “Analysis of QoS requirements for e-health services and mapping to evolved packet system QoS classes,” *Int. J. Telemedicine Appl.*, p. 9:18, Jan. 2010.
- [16] S. Soro and W. Heinzelman, “A survey of visual sensor networks,” *Advances in Multimedia*, pp. 1–22, 2009.
- [17] A. Mavrinac and X. Chen, “Modeling coverage in camera networks: A survey,” *Int. Jour. of Comp. Vision*, vol. 101, no. 1, pp. 205–226, 2013.
- [18] S. Hengstler, D. Prashanth, S. Fong, and H. Aghajan, “Mesheye: a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance,” in *Proc. of the 6th International Symposium on Information Processing in Sensor Networks (IPSN '07)*. Cambridge, Massachusetts, USA: ACM Press, 2007, pp. 360–369.
- [19] Z. Yang and K. Nahrstedt, “A bandwidth management framework for wireless camera array,” in *Proc. of the international workshop on Network and operating systems support for digital audio and video (NOSSDAV '05)*, Stevenson, Washington, USA, 2005, pp. 147–152.
- [20] A. Ercan, D. Yang, A. E. Gamal, and L. Guibas, “Optimal placement and selection of camera network nodes for target localization,” in *Distributed Computing in Sensor Systems*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, pp. 389–404 vol. 4026.
- [21] W. Wolf, B. Ozer, and T. Lv, “Smart cameras as embedded systems,” *Computer*, vol. 35, pp. 48–53, 2002.
- [22] K. Morioka and H. Hashimoto, “Appearance based object identification for distributed vision sensors in intelligent space,” in *Proc. of Intelligent Robots and Systems (IROS '04). IEEE/RSJ International Conference on*, vol. 1, Sendai, Japan, 2004, pp. 199–204.

- [23] K. Obraczka, R. Manduchi, and J. Garcia-Luna-Aveces, “Managing the information flow in visual sensor networks,” in *Proc. of the 5th Wireless Personal Multimedia Communications (WPMC '02), International Symposium on*, vol. 3, Sheraton Waikiki, Honolulu, Hawaii, 2002, pp. 1177–1181.
- [24] T. H. Ko and N. M. Berry., “On scaling distributed low-power wireless image sensors,” in *Proc. of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06)*, vol. 9, Washington, DC, USA, 2006, p. 235.
- [25] N. Zamora and R. Marculescu, “Coordinated distributed power management with video sensor networks: Analysis, simulation, and prototyping,” in *Proc. of the Distributed Smart Cameras (ICDSC '07). 1th ACM/IEEE International Conference on*, 2007, pp. 4–11.
- [26] S. Soro and W. Heinzelman, “Camera selection in visual sensor networks,” in *Proc of the Advanced Video and Signal Based Surveillance (AVSS '07). IEEE Conference on*, 2007, pp. 81–86.
- [27] D. Yang, J. Shin, A. O. Ercan, and L. Guibas, “Sensor tasking for occupancy reasoning in a network of cameras,” in *Proc. of the IEEE/ICST 1st Workshop on Broadband Advanced Sensor Networks (BASENETS '04)*, San Josè, CA, USA, 2004.
- [28] J. Park, P. Bhat, and A. Kak, “A look-up table based approach for solving the camera selection problem in large camera networks,” in *Proc. of the International Workshop on Distributed Smart Cameras (DCS '06)*, 2006.
- [29] D. Lindberg, “The theory of pinhole images from antiquity to the thirteenth century,” *Archive for History of Exact Sciences*, vol. 5, no. 2, pp. 154–176, 1968.
- [30] D. Dong and J. Atick, “Statistics of natural time-varying images,” *Network: Computation in Neural Systems*, vol. 6, no. 3, pp. 345–358, 1995.
- [31] J. Bouguet, “Camera calibration toolbox for matlab,” 2004. [Online]. Available: <http://www.vision.caltech.edu>
- [32] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library.* ” O'Reilly Media, Inc.”, 2008.
- [33] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Image Analysis.* Springer, 2003, pp. 363–370.

- [34] P. ITU-T RECOMMENDATION, “910 subjective video quality assessment methods for multimedia applications,” 2008. [Online]. Available: <http://handle.itu.int/11.1002/1000/9317>
- [35] J. Lee, F. D. Simone, and T. Ebrahimi, “Subjective quality evaluation via paired comparison: application to scalable video coding,” *Multimedia, IEEE Transactions on*, vol. 13, no. 5, pp. 882–893, 2011.
- [36] J. Kemeny, “Mathematics without numbers,” *Daedalus*, vol. 88, no. 4, pp. 577–591, 1959.
- [37] H. P. Young and A. Levenglick, “A consistent extension of condorcet’s election principle,” *SIAM Journal on Applied Mathematics*, vol. 35, no. 2, pp. 285–300, 1978.
- [38] C. Spearman, “The proof and measurement of association between two things,” *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
- [39] M. Kendall and J. Gibbons, *Rank correlation methods*, ser. A Charles Griffin Book. E. Arnold, 1990.
- [40] R. Kumar and S. Vassilvitskii, “Generalized distances between rankings,” in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 571–580.
- [41] S. Akter, P. Ray, and J. D’Ambra, “Continuance of mhealth services at the bottom of the pyramid: the roles of service quality and trust,” *Electronic Markets*, vol. 23, no. 1, pp. 29–47, 2013.
- [42] G. Blumrosen, N. Avisdris, R. Kupfer, and B. Rubinsky, “C-smart: Efficient seamless cellular phone based patient monitoring system,” in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1–6.
- [43] S. Mougiakakou, I. Kouris, D. Iliopoulou, A. Vazeou, and D. Koutsouris, “Mobile technology to empower people with diabetes mellitus: design and development of a mobile application,” in *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*. IEEE, 2009, pp. 1–4.
- [44] E. Lai and K. Friedl, “Digital soldiers: Transforming personalized health in challenging and changing environments,” in *Wearable Micro and Nano Technologies for Personalized Health (pHealth), 2009 6th International Workshop on*. IEEE, 2009, pp. 5–8.

- [45] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, “Codeblue: An ad hoc sensor network infrastructure for emergency medical care,” in *International workshop on wearable and implantable body sensor networks*, vol. 5, 2004.
- [46] I. Korhonen, J. Pärkkä, and M. V. Gils, “Health monitoring in the home of the future,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 22, no. 3, pp. 66–73, 2003.
- [47] C. Delgorge, F. Courrèges, L. Bassit, C. Novales, C. Rosenberger, N. Smith-Guerin, C. Brù, R. Gilabert, M. Vannoni, G. Poisson *et al.*, “A tele-operated mobile ultrasound scanner using a light-weight robot,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 9, no. 1, pp. 50–58, 2005.
- [48] Cisco, “Global mobile data traffic forecast update, 2011-2016,” Cisco VNI White Paper, Tech. Rep., Feb 2012.
- [49] Z. Fan *et al.*, “M2m communications for e-health: Standards, enabling technologies, and research challenges,” in *2012 6th International Symposium on Medical Information and Communication Technology (ISMICT)*, 2012, pp. 1–4.
- [50] 3GPP TS 36.300, “Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2,” version 10.8.0 Release 10, Tech. Rep., 2012.
- [51] 3GPP TS 23.203, “Policy and charging control architecture,” version 10.7.0 Release 10, Tech. Rep., 2012.
- [52] H. Myung, J. Lim, and D. Goodman, “Single carrier FDMA for uplink wireless transmission,” *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30–38, 2006.
- [53] M. Rumney, “3GPP LTE: Introducing single-carrier FDMA,” *Agilent measurement journal*, vol. 4, pp. 18–27, 2008.
- [54] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the h. 264/avc standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [55] K. Tappayuthpijarn, G. Liebl, T. Stockhammer, and E. Steinbach, “Adaptive video streaming over a mobile network with TCP-friendly rate control,” in *The 2009 International Conference on Wireless Communications and Mobile Computing (IWCMC’09)*, Leipzig, Germany, 2009, pp. 1325–1329.

- [56] I. Kofler, J. Seidl, C. Timmerer, H. Hellwagner, I. Djama, and T. Ahmed, “Using MPEG-21 for cross-layer multimedia content adaptation,” *Springer Journal on Signal, Image and Video Processing*, vol. 2, pp. 355–370, 2008.
- [57] T. Pliakas, G. Kormentzas, and S. Tsekeridou, “Joint scalable video coding and packet prioritisation for video streaming over IP/802.11e heterogeneous networks,” in *Proceedings of the 3rd International ICST Mobile Multimedia Communications Conference*, Nafpakos, Greece, August 2007, paper 31.
- [58] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer, “Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG-4-AVC,” in *IEEE International Conference on Multimedia and Expo (ICME’06)*, Toronto, Canada, 2006, pp. 625–628.
- [59] H. Juan, H. Huang, C. Huang, and T. Chiang, “Cross-layer mobile WiMAX MAC designs for the H.264/AVC scalable video coding,” *Springer Wireless Networks*, vol. 16, pp. 113–123, January 2010.
- [60] G. Bianchi, A. Detti, P. Loreti, C. Pisa, F. Proto, W. Kellerer, S. Thakolsri, and J. Widmer, “Application-aware H.264 scalable video coding delivery over wireless LAN: experimental assessment,” in *Second International Workshop on Cross Layer Design (IWCLD’09)*, Palma de Mallorca, Spain, 2009, pp. 1–6.
- [61] Y. Fallah, P. Nasiopoulos, and H. Alnuweiri, “Scheduled and contention access transmission of partitioned H.264 video over WLANs,” in *IEEE Global Communications Conference*, Washington DC, USA, 2007, pp. 2134–2139.
- [62] IEEE, “MAC enhancements for robust audio video streaming,” 2012, IEEE Std 802.11aa-2012.
- [63] T. Sutinen, J. Vehkaperä, E. Piri, and M. Uitto, “Towards ubiquitous video services through scalable video coding and cross-layer optimization,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, 2012.
- [64] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos, “Channel aware multiuser scalable video streaming over lossy under-provisioned channels: Modeling and analysis,” *IEEE Trans. on Multimedia*, vol. 10, no. 7, pp. 1366–1381, 2008.
- [65] S. Cicalò, A. Haseeb, and V. Tralli, “Fairness-oriented multi-stream rate adaptation using scalable video coding,” *Elsevier Signal Processing: Image Communication*, 2012.

- [66] Y. Wang, L. Chau, and K. Yap, “Joint rate allocation for multiprogram video coding using fgs,” vol. 20, no. 6, pp. 829–837, 2010.
- [67] X. Zhang, A. Vetro, Y. Shi, and S. Huifang, “Constant quality constrained rate allocation for fgs-coded video,” vol. 13, no. 2, pp. 121–130, 2003.
- [68] M. Dai, D. Loguinov, and H. Radha, “Rate-distortion analysis and quality control in scalable internet streaming,” vol. 8, no. 6, pp. 1135–1146, 2006.
- [69] D. Kwon, M. Shen, and C. Kuo, “Rate control for h.264 video with enhanced rate and distortion models,” vol. 17, no. 5, pp. 517–529, 2007.
- [70] A. Haseeb, M. Martini, S. Cicalò, and V. Tralli, “Rate and distortion modeling for real-time MGS coding and adaptation,” in *Wireless Advanced (WiAd), 2012*, june 2012, pp. 85–89.
- [71] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels,” *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 6, pp. 1012–1032, 2000.
- [72] C. Castellanos, D. Villa, C. Rosa, K. Pedersen, F. Calabrese, P. Michaelson, J. Michel *et al.*, “Performance of uplink fractional power control in UTRAN LTE,” in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*. IEEE, 2008, pp. 2517–2521.
- [73] I. Wong, O. Oteri, and W. McCoy, “Optimal resource allocation in uplink sc-fdma systems,” *Wireless Communications, IEEE Transactions on*, vol. 8, no. 5, pp. 2161–2165, 2009.
- [74] JSVM, “9.19.11 reference software,” Tech. Rep., 2011. [Online]. Available: <http://www.hhi.fraunhofer.de/departments/video-coding-analytics>
- [75] M. Razaak, M. Martini, and K. Savino, “A study on quality assessment for medical ultrasound video compressed via HEVC,” *to appear, IEEE Journal on Biomedical Informatics, 2014*.
- [76] R. Ferrus, O. Sallent, G. Baldini, and L. Goratti, “LTE: the technology driver for future public safety communications,” *Communications Magazine, IEEE*, vol. 51, no. 10, pp. 154–161, Oct 2013.
- [77] J. Gallego, A. Hernandez-Solana, M. Canales, J. Lafuente, A. Valdovinos, and J. Fernandez-Navajas, “Performance analysis of multiplexed medical data transmission for mobile emergency care over the umts channel,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 13–22, March 2005.

- [78] C. Doukas and I. Maglogiannis, “Adaptive transmission of medical image and video using scalable coding and context-aware wireless medical networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, no. 1, 2008.
- [79] D. Shaeffer, “Mems inertial sensors: A tutorial overview,” *Communications Magazine, IEEE*, vol. 51, no. 4, pp. 100–109, Apr 2013.
- [80] D. Dardari, P. Closas, and P. Djuric, “Indoor tracking: Theory, methods, and technologies,” *Vehicular Technology, IEEE Transactions on*, vol. 64, no. 4, pp. 1263–1278, Apr. 2015.
- [81] B. Sobhani, E. Paolini, A. Giorgetti, M. .Mazzotti, and M. Chiani, “Target tracking for UWB multistatic radar sensor networks,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 1, pp. 125–136, Feb. 2014.
- [82] A. Giorgetti and M. Chiani, “Time-of-arrival estimation based on information theoretic criteria,” vol. 61, no. 8, pp. 1869–1879, 2013.
- [83] G. Glanzer, “Personal and first-responder positioning: State of the art and future trends,” in *Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS), 2012*, Oct 2012, pp. 1–7.
- [84] M. Z. Win, A. Conti, S. Mazuelas, Y. Shen, W. M. Gifford, D. Dardari, and M. Chiani, “Network localization and navigation via cooperation,” vol. 49, no. 5, pp. 56–62, May 2011.
- [85] E. Paolini, A. Giorgetti, M. Chiani, R. Minutolo, and M. Montanari, “Localisation capability of cooperative anti-intruder Radar systems,” vol. 2008, pp. 1–14, Jun. 2008.
- [86] S. Severi, G. Liva, M. Chiani, and D. Dardari, “A new low-complexity user tracking algorithm for WLAN-based positioning systems,” in *Proc. of 16th IST Mobile & Wireless Communication Summit*, Budapest, Hungary, Jul. 2007.
- [87] S. Cicalò and V. Tralli, “Adaptive resource allocation with proportional rate constraints for uplink SC-FDMA systems,” *IEEE Communications Letters*, vol. 18, no. 8, pp. 1419–1422, Aug 2014.
- [88] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

- [89] S. S. Cicalò, N. Changuel, R. Miller, B. Sayadi, and V. Tralli, “Quality-fair adaptive streaming over LTE networks,” in *in Proc.of IEEE 39th International conference on Acoustic, Speech and Signal Processing*, May 2014, pp. 1–5.
- [90] S. Cicalò and V. Tralli, “Distortion-fair cross-layer resource allocation for scalable video transmission in OFDMA wireless networks,” *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 848–863, April 2014.
- [91] G. T. 36.104, “Evolved universal terrestrial radio access (E-UTRA); base station radio transmission and reception,” version 9.5.0 Release 10, Tech. Rep., 2010.
- [92] A. Varga and R. Hornig, “An overview of the OMNet++ simulation environment,” in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops*, ser. Simutools ’08. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 60:1–60:10.
- [93] P. Series, “Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 mhz to 100 ghz,” 2009.
- [94] C. Danae, “Technical specification group radio access network,” Technical report, Technical report, Spreading and Modulation, <http://www.3gpp.org>, Tech. Rep., 1999.

# Acknowledgement

I am using this final page to express my sincere gratitude to everyone who supported me throughout the achievement of my doctoral degree.

Firstly, I would like to thank my advisor, Prof. Marco Chiani, for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge.

A special thank you goes to Matteo Mazzotti: his guidance helped me in all the time of research. I could not have imagined having a better colleague and mentor for my Ph.D. career. We have shared many adventures together collecting successes and professional satisfactions and now I am really honoured to be one of your friends.

I would like to thank all the telecommunications group guys at the University of Bologna, Cesena campus, for making this a wonderful working place, starting from my Ph.D. fellows, Francesco, Nicolás, Andrea, Ahmed, Ashraf, Emanuele and Vincenzo, and also Enrico, Andrea Giorgetti and Prof. Davide Dardari. A special thank to Anna for her help, support and for all we shared in these three years. Again, I would like to thank all the APICe colleagues and in particular Danilo, Alex and Roberto.

I would like to thank Prof. Tommaso Melodia, who allowed me to visit him at the Northeastern University in Boston and for the fruitful research collaboration that we are carrying on. Thanks to whom made me feel welcome in Boston: Enrico, Nancy, Guan, Emreacan and Liyang.

Dedico questa mia vittoria a tutte quelle persone che mi accompagnano e sostengono quotidianamente.

Il mio ringraziamento va alla mia famiglia: a mio padre Luigi, a mia madre Mara, a mio zio Massimo e a mia nonna Lora (so che mi puoi sentire anche da lassù): sono loro che mi hanno insegnato cosa significhi il sacrificio, il rispetto, l'educazione e l'affetto. Sarete sempre i pilastri fondamentali della mia vita.

Ringrazio Lorenzo e Tony, due dei miei piú cari amici, la cui vicinanza é sempre stata fondamentale; grazie inoltre ai miei due fedeli compagni di avventure universitarie: Alex e Michele. Vi voglio bene. Grazie di cuore a due amiche speciali come Natasha e Vanessa: siete la dimostrazione che nessuna distanza puó dissolvere una amicizia cosí profonda.

Un ringraziamento particolare a Monica a cui non devo essere grato solo per l'amicizia fraterna, ma soprattutto per gli insegnamenti e le lezioni di vita che porteró sempre con me, ovunque andró. Grazie.

Manca solo una persona, speciale, all'appello. Una persona con cui ho condiviso dei momenti unici e che mi ha insegnato tante cose (anche se non ci crederai é cosí). Ti voglio bene Lisa. "There are no goodbyes for us. Wherever you are...", ricordatelo sempre.

BUONA VITA A TUTTI!