Multiple Video Delivery in m-Health Emergency Applications

Sergio Cicalò, Matteo Mazzotti, Simone Moretti, Velio Tralli, and Marco Chiani

***Abstract*—M-health services are expected to become increasingly relevant in the management of emergency situations by enabling real-time support of remote medical experts. In this context, the transmission of multiple health-related video streams from an ambulance to a remote hospital 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. In order to deliver the multiple streams over a single bandwidth-limited wireless access channel, in this paper we propose a novel optimization framework that enables to classify the available video sources and to automatically select and adapt the best streams to transmit. The camera ranking algorithm jointly works with a cross-layer adaptation strategy for multiple scalable streams to achieve different objectives and/or tradeoffs in terms of number and target quality of the transmitted videos. The final goal of the optimization is to dynamically adjust the overall transmitted throughput to meet the actual available bandwidth, while being able to provide high quality to diagnostic video sequences and lower quality to less critical ambient videos. Numerical simulations considering a realistic emergency scenario with LTE-Advanced connectivity show that the proposed content/context-aware solution is able to automatically select the best sources of information from a visual point of view and to achieve optimal end-to-end video quality for both the diagnostic and the ambient videos.**

I. Introduction

Nowadays e-health is one of the most promising applications of emerging information and communication technologies [1]. In particular, tele-medicine services can highly benefit from the recent advances offered by mobile communication systems [2], which are now potentially able to support a wide range of ubiquitous health-care applications, such as tele-diagnosis [3], real-time monitoring of vital parameters [4], remote treatment of patients [5] and even tele-surgery.

In tele-medicine, the reliable transmission of heterogeneous health-related information is being increasingly used for remote patient monitoring and disease management. In this regard, the usage of a visual sensor network (VSN) can be employed for remote monitoring applications where video information is useful for controlling and managing particular events [6]. In emergency situation, the delivery of video information to a remote hospital allows the hospital personnel to support and manage the first-aid operations especially when the number of operators on the field is limited. Moreover, the specialists may be allowed to perform preliminary diagnostic analysis and to prepare the hospitalization adequately. The presence of a 4G access network, *e.g.*, Long Term Evolution Advanced (LTE-A), can be exploited to establish a communication link with the emergency area, as done, *e.g.*, in the recent proposals for next-generation public safety networks [7], [8].

However, despite the recent enhancements in the 4G cellular networks, one of the most critical issue concerning multiple video transmission still consists in the large amount of data that each camera collects from the monitored environment. For this reason, in the context of redundant camera networks, source selection and management play a fundamental role. Each camera is characterized by a directional sensing model: the collected information depends on the direction in which the camera is oriented and the 3D viewing volume defined by the camera field of view (FoV). In [9] a camera selection algorithm is provided with the aim to reconstruct a view from a user-specified view point. Other works address the specific problem of camera placement. An example is reported in [10], where a smart camera-network deployment is proposed to reach a given coverage objective, while in [11] a multi-camera management model is described, based on a collaborative scene analysis aiming to evaluate the area occupancy. In this work we assume that the camera network is already deployed across the emergency area: some of the devices are stationary, while others might be moving, e.g. as a part of the equipment worn by the first-aid responders. In each moment a subset of cameras has to be selected in order to perform continuous monitoring with the required quality.

The selected videos (ambient and/or diagnostic) are collected and aggregated to coordinate the transmissions to and from the hospital [12]. The possibility to adapt the video encoding to the current transmission conditions becomes particularly important in the context of emergency m-health applications, because mobile links are error-prone and are usually characterized by limited and variable bandwidth [5][13]. For these reasons, several works have considered optimized cross-layer methods to improve the reliability and the quality of m-health video applications [14]–[22].

In [14] a robust cross-layer method to deliver real-time robotic ultrasonography video over 3G/WLAN network is presented. The proposed strategy controls both video encoding and physical layer parameters to minimize the total distortion with a total rate constraint. Numerical results considering a Wireless LAN link show that improvement in terms of both objective medical video quality and subjective quality is achieved by using the proposed system. In [15] an adaptive rate control algorithm has been proposed for H.264/AVC video streaming in bandwidth demanding m-health applications over

S. Cicalò and V. Tralli are with CNIT, Department of Engineering, University of Ferrara, Italy. e-mail: {sergio.cicalo, velio.tralli}@unife.it. M. Mazzotti is with OCEM Airfield Technology, Crespellano (BO), Italy. e-mail: matteo.mazzotti@ocem.com. S. Moretti and M. Chiani are with CNIT, Department of Electrical, Electronic, and Information Engineering (DEI), University of Bologna, Italy. e-mail: {simone.moretti6, marco.chiani}@unibo.it.

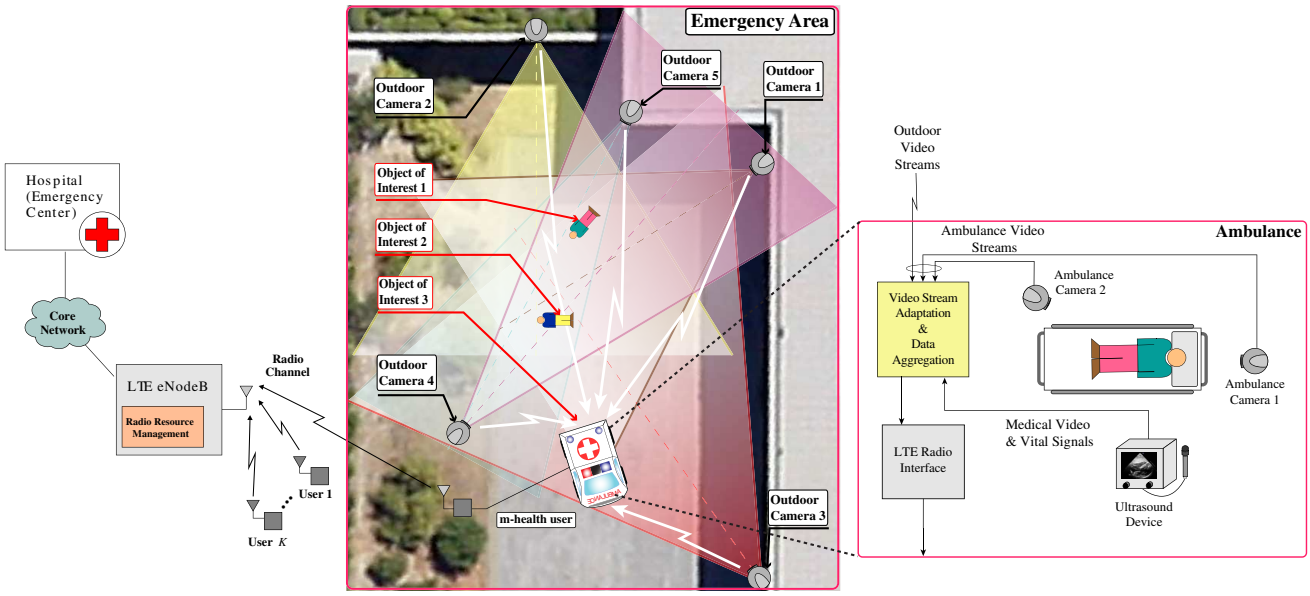


Fig. 1. The proposed m-health architecture for emergency scenarios.

beyond-3G/4G networks. The approach is intended to satisfy some pre-defined medical QoS requirements in terms of video quality, frame-rate and end-to-end delay. The proposed scheme tested in a HSDPA testbed has shown a better capability to dynamically satisfy the medical QoS requirements with respect to the standard H.264 rate control. In [17] this framework has also been tested in HSUPA and mobile WiMAX networks. Wireless transmission of H.264/AVC medical ultrasound video over mobile WiMAX networks is also the focus of the paper in [19]. Through extensive subjective evaluations resulting from a realistic testbed, the Authors checked the ability of 4G networks to provide robust and clinically acceptable high-resolution ultrasound video transmission.

The content/context-aware optimization of m-health video streaming in emergency situation has been recently addressed in [20], [21] and [22]. The Authors in [20] and [21] have proposed a full cross-layer optimized architecture with integrated building blocks for medical media content fusion, delivery and access, even on the move in emergency contexts. The novel system architecture has been successfully evaluated by both simulations and medical staff evaluations. The Authors in [22] have proposed an m-health video communication system for emergency telemedicine and response in disaster incidents. Numerical results involving H.265/HEVC standard and realistic modeling of a Wireless LAN network have shown that a proper trade-off can be found in terms of encoding time, bitrate, and video quality of the video to transmit.

All of the aforementioned works consider a single m-health uplink video transmission and single-layer encoding schemes, *e.g.*, H.264/AVC or H.265/HEVC. The latter provide high coding efficiency but lacks from flexibility in rate adaptation. In this work we consider the scalable extension of the H.264/AVC standard, known as scalable video coding (SVC) [23], which conjugates good compression efficiency and wide flexibility for the adaptation to the wireless channel conditions. In fact, several solutions have been recently proposed for e-health applications based on SVC, *e.g.*, [24]. The content-aware optimization of multiple SVC video transmission in the downlink of cellular network has been extensively investigated in the literature, *e.g.*, in [25], [26], where generally a proxy is in charge of optimizing the multiple video streams delivered to a set of heterogeneous users according to different QoE/QoS constraints. However, to the best of Authors' knowledge, the joint context- and content-aware optimization of the m-health SVC multiple video streaming has not yet been considered.

In this paper we propose a novel optimization framework in order to automatically select and adapt the best video streams to transmit over a single bandwidth-limited wireless access channel. We first introduce a camera ranking technique for the VSN deployed in the emergency area. It permits to select one or more cameras taking into account specific quality criteria mainly related to the visual representation of the object of interest. Then, we describe a novel solution for the transmission of the videos from the emergency area, based on the joint selection, adaptation and aggregation of the streams directly performed at the application layer of the processing equipment inside the ambulance. 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.

In our solution the adaptation exploits the information on the camera visual qualities and on the available rate assigned by the LTE-A wireless access network in order to optimize the quality and/or fairness according to two proposed strategies. The first strategy aims at delivering the highest visual and video qualities according to different priorities, derived from the camera ranking outcomes, and some suitably defined fairness constraints. The second strategy aims at maximizing the aggregated weighted quality, where the weights are derived

according to the visual ranking process. Both strategies are based on the modeling and evaluation of objective video quality metrics and are tested and compared in the considered emergency scenarios.

It is shown that the proposed framework permits to achieve the optimal end-to-end quality even in the presence of rate limitations and fluctuations due to the wireless channel and the intense traffic within the LTE-A cell. When the channel conditions deteriorate, the adaptation strategy appropriately discards the cameras which provide the lowest visual quality based on the camera ranking results, thereby ensuring the real-time transmission of the medical video stream at the required diagnostic quality.

The framework proposed in this paper is built on the preliminary contribution in [27], where we introduced the basic ideas behind the camera ranking algorithm (CRA) and the fairness-oriented adaptation strategy, but without considering their joint design and optimization and the novel efficiency-oriented adaptation strategy.

The paper is organized as follows. The overall system architecture is described in Section II. In Section III we present the camera ranking strategy, while Section IV and V detail the SVC quality-to-rate model and the video adaptation strategies, respectively. Finally, in Section VI, we provide several numerical results concerning a simulated emergency scenario where a camera network is deployed with monitoring purposes.

II. System Architecture

The m-health scenario addressed in this paper is depicted in Fig. 1. 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, we assume that the cameras' position and attitude, *i.e.*, the angle between the optical axis and the direction of the camera's motion, are known in real-time, *e.g.*, by means of inertial units mounted on the devices [28], GPS receivers or other real-time locating system (RTLS) [29][30]. Modern equipment for emergency teams, in fact, are nowadays more frequently capable to provide location information about people and objects on the field, in order to improve the efficiency and the safety of the operations [31].

Multimedia flows are collected by the ambulance, processed in real-time and possibly 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 Long Term Evolution (LTE)/evolved packet core (EPC) network. LTE allows to prioritize health-care related video traffic over less

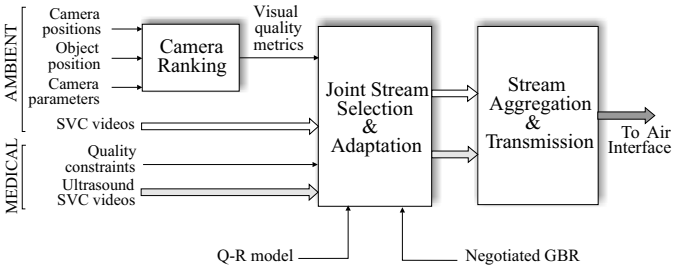


Fig. 2. Block diagram reporting the exchange of information in the proposed multi-camera management system.

critical traffic. In particular, LTE provides a standardized traffic class differentiation based on some sets of predefined QoS requirements related to throughput, delay and packet loss rate [32]. Each traffic class is identified by a scalar value, called quality class indicator (QCI). The QCI classes fall within two macro-categories: guaranteed bit rate (GBR) and non-GBR. An example of a mapping schemes of health-related information to LTE traffic classes can be found in [8].

We consider that 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 (non-GBR) users, indexed by the sets $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively. The e-NodeB tries to guarantee the transmission rates R_0 to the m-health user[1] and 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 remaining capacity available after allocating all GBR users [33] [34].

In this work, we assume 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 phase a patient is loaded on the ambulance, where an ultrasound examination is performed to check his/her health status. During this operation, the ambulance collects and sends up to two ambient indoor videos and one diagnostic video sequence. In both situations, the selected videos are sent through the available LTE-A 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 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. 2 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 CRA processes the side information including the positions and the parameters of the ambient cameras and provides as output

[1]Note that here, as well as in the rest of the paper, we indicate the m-health GBR user with the subscript 0.

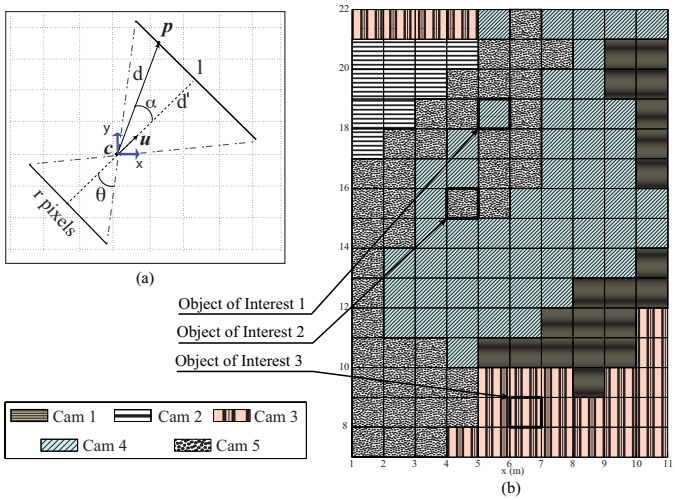


Fig. 3. (a) the 2D pinhole camera model. In evidence the FoV, the camera orientation and the image linear dimension, l, corresponding to r pixels on the camera. (b) the camera ranking algorithm results for the whole emergency area.

a set of visual quality metrics. Based on these values, the following unit selects the optimal set of ambient videos to transmit, while guaranteeing a sufficient video quality for high priority diagnostic streams (when present). As discussed in the following sections, real-time video selection and adaptation are jointly carried out.

III. Multi camera ranking

In this section we describe a new CRA, based on an ordering criterion dependent on the quality of representation of the object of interest. According to Fig. 1, we assume that the area is monitored through a multi-camera network connected to the ambulance through local wireless links. The proposed ranking technique is based on geometrical considerations, as we assume 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 fields of view (FoVs). Based on camera FoV and spatial orientation, it is possible to determine which portion of the area of interest is monitored 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. For an object at distance d from the camera, the visual quality is measured here as the amount of pixels required to represent a surface of unit area, orthogonal to the camera orientation, placed at distance d from the camera.

To describe the video acquisition devices we adopted a pinhole camera model, as depicted in Fig. 3a. With reference to a generic element of the VSN, dash-dotted lines represent one of the FoVs, the optical center is identified by the point $\mathbf{c}$, while the target to be monitored is pointed by $\mathbf{p}$. The camera orientation is described by the unit norm vector $\mathbf{u}$, *i.e.*, with $\|\mathbf{u}\| = 1$. For the sake of simplicity and without loss of generality (*w.l.o.g.*), we assume that the optical center of the considered camera is positioned in the origin of the reference system. With this assumption, the number of linear pixels per meter required to represent the target in $\boldsymbol{p}$ can be calculated as

$$q_{\text{lin}} = \frac{r}{2\,\boldsymbol{p}\cdot\boldsymbol{u}\tan\theta} \qquad [ppm] \tag{1}$$

where 2θ is the width of the FoV. Given the camera resolution, r represents the number of linear pixels along the considered image dimension.

Clearly, assuming a squared aspect ratio for the pixels, the number of pixels per unit area at the considered distance is $q = q_{\text{lin}}^2$, and it numerically represents the *visual quality* provided by the considered camera. We remark that the proposed ranking algorithm is intended for static or low-motion scenes. Novel camera selection strategies taking into account acquisition frame rate and higher velocity for the elements in the scene are currently object of study.

To illustrate how the CRA works, we consider the complete scenario depicted in Fig. 1. The emergency area (monitored by 5 cameras) has been partitioned in sub-areas, *i.e.*, 1 m $\times$ 1 m square surfaces, as depicted in Fig. 3b. For each sub-area, the proposed algorithm ranks the available cameras based on the visual quality. In Fig. 3b, we show the camera providing the best visual quality 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 (corresponding to an injured person on the ground in our simulation) 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 actual application the position of the object of interest for the emergency, *i.e.*, the square in the grid, is selected by the operator.

In the following we will denote with the ranking value q_v the visual quality of the v-th camera. To simplify the notation, we assume *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 $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots \subset \mathcal{V}_N$. For Reader's convenience, the most used symbols and acronyms of this paper are summarized in Table I.

IV. SVC: Quality-Rate Model

The SVC standard is attractive for applications 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. In this paper we focus on video quality scalability, also known as signal to noise ratio

Notation	Description	Notation	Description
$\mathcal{K}$, K	Set and total number of users	N	Total number of cameras
R_k	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^{target}	Target utility of video v	U_v	Utility of video v
F_v^{bl}	BL rate for video v	F_v^{hl}	Maximum source rate for video v
α	Shaping value	H	Overhead factor
Acronym	**Full name**	**Acronym**	**Full name**
LTE	Long Term Evolution	BS	base station
SVC	scalable video coding	Q-R	Quality-rate
SSIM	structural similarity	GOP	group of pictures
BL	base layer	EL	enhancement layer
GBR	guaranteed bit rate	SNR	signal to noise ratio
IDR	intra decoding refresh	CRA	camera ranking algorithm

TABLE I
LIST OF OF MOST USED SYMBOLS AND ACRONYMS

(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 [23], namely the coarse grain scalability (CGS) and the medium grain scalability (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 paper we focus on MGS scalability.

QoE-aware 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. In order to allow for real-time content-aware adaptation, the video quality can be evaluated according to a particular objective video quality metric, *e.g.*, mean square error (MSE), peak signal to noise ratio (PSNR) or structural similarity (SSIM) metrics [35]. Here, we consider the SSIM 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 [36]. 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 pictures (GOPs). The interval between two consequent I-frames, also known as intra decoding refresh (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 [37] and recently extended in [26], [38], 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 $F \in [F_v^{\text{bl}}, F_v^{\text{hl}}]$, where F_v^{hl} and F_v^{bl} represent the two limit rates corresponding to all the available enhancement layers (ELs) and the sole base layer (BL), 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 BL.

Although the framework we propose holds for any monotonic strictly increasing and invertible function $U_v(F)$, in this paper we consider the following parametric utility model [38]:

$$U_v(F) = a_v \ln(F - b_v) + c_v, \quad a_v > 0, b_v \leq F_v^{\text{bl}} \qquad (2)$$

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 (2). 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 Figure 4, an example of the empirical Q-R relationship for a randomly chosen set of pictures in a IDR period is shown for six ambient video sequences and one ultrasonography video. The related Q-R curves based on model (2) are also reported for comparison. The encoding parameters are reported in Table II and III. As noted in the figure, larger coding rates have been considered for the ultrasound sequence, since diagnostic videos often require very high video quality in order to permit accurate diagnosis by remote specialists [36][39].

V. JOINT VIDEO SELECTION AND SVC ADAPTATION

We assume that the ambulance equipment negotiates with the LTE access network a GBR 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. We consider here two different strategies with the goal of jointly selecting the best set of videos to be transmitted and adapting the streams to the available channel. The first strategy aims at delivering the highest visual and video qualities

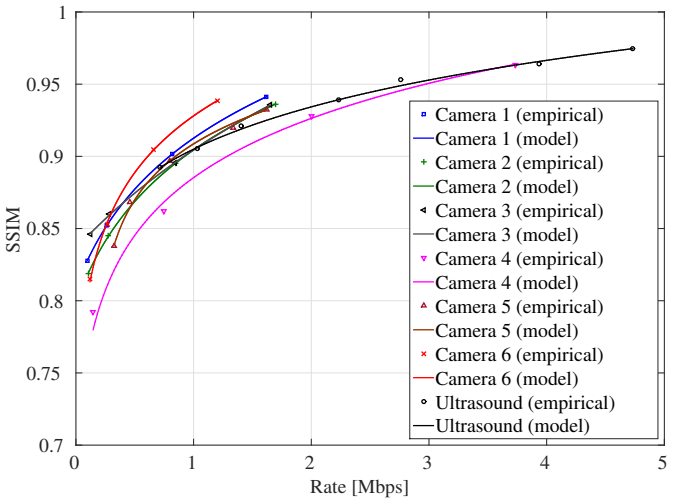


Fig. 4. Utility models in terms of the SSIM.

according to some fairness constraints resulting from the CRA outcomes. The second strategy aims at maximizing the aggregated weighted utility, where the weights are derived from the CRA. 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 and are described in the following subsections V-A and V-B, respectively.

A. 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 [26], addressing the problem for both empirical and semi-analytical rate-distortion (R-D) models, and then extended in [33] to consider the different priorities of the video sequences. However, the Authors in [26], [33] 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. Since the video sources work with sets of discrete values of rates, the optimization problem in its primitive form, introduced in [26], can be relaxed into the following optimization problem based on continuous values of rates

$$\max_{1\leq x\leq N}\max_{F_v, v\in\mathcal{V}_x}\sum_{v\in\mathcal{V}_x} w_v U_v(F_v) \tag{3a}$$

$$s.t.\ U_v \geq U_v^{\text{target}}, \qquad \forall v\in\mathcal{V}_x \tag{3b}$$

$$\sum_{v\in\mathcal{V}_x} HF_v \leq R_0, \tag{3c}$$

$$F_v^{\text{bl}} < F_v < F_v^{\text{hl}}, \qquad \forall v\in\mathcal{V}_x \tag{3d}$$

$$\Delta_{\boldsymbol{p}}(U_i,U_j)=0, \qquad \forall i,j\in\mathcal{V}_x, \tag{3e}$$

where H is the estimated overhead introduced at the different layers of the network architecture, F_v and U_v^{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 (3c) and (3d) impose that the adapted video sequences must be supported in real-time by the available transmission GBR R_0, thereby allocating at least the BL rate and no more than the maximum encoding rate to each selected video stream.

The set of conditions in (3e) is verified when the quality difference among the different videos is minimized with continuous rates. The function $\Delta_{\boldsymbol{p}}(U_i,U_j)$ is the utility-fairness metric for each pair of videos. It extends the fairness metric introduced in [26] to include priority information. It is defined as

$$\Delta_{\boldsymbol{p}}(U_i,U_j)=\begin{cases}0 & \text{if } (i,j)\in\mathcal{F}\vee(j,i)\in\mathcal{F}\\ \delta_{\boldsymbol{p}}(U_i,U_j) & \text{otherwise}\end{cases} \tag{4}$$

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

$$\begin{aligned}\mathcal{F}=\{(i,j):&(U_i=U_i^{\min}\wedge p_i(1-U_i)>p_j(1-U_j))\vee\\ &(U_i=U_i^{\text{hl}}\wedge p_i(1-U_i)<p_j(1-U_j))\}\end{aligned} \tag{5}$$

where $\boldsymbol{p}=[p_1,\ldots,p_N]\succeq\mathbf{0}$ 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 (4) extends the elementary fairness metric $\delta_{\boldsymbol{p}}(U_i,U_j)$ 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 amount of 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 for reasonable values of GBR an SSIM 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. We remark that, as the GBR decreases, the difference in quality between macro-classes will also decreases and *vice versa*.

Finally, the weights w_v in (3a) are used to account for the different visual qualities according to the outcome of the CRA, and are evaluated as follows:

$$w_v=\left(\frac{q_v}{\sum_{s\in\mathcal{V}_N}q_s}\right)^{\alpha},\forall v\in\mathcal{V}_x \tag{6}$$

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 small value of α favors the transmission of more videos with lower quality,

Algorithm 1 Pseudo code to solve problem (3)

Input: $\mathcal{V}$, R_0; $F_v^{\text{bl}}, F_v^{\text{hl}}, U_v^{\text{target}}, w_v, a_v, b_v, c_v, \forall v \in \mathcal{V}_N$;
for all $x = N, \dots, 1$ **do**
  if $\sum_{v\in\mathcal{V}_x} HF_v^{\min} \le R_0$ **then**
    if $\sum_{v\in\mathcal{V}_x} HF_v^{\text{hl}} \le R_0$ **then**
      set $\tilde{F}_v^{(x)} = F_v^{\text{hl}}$, $\forall v \in \mathcal{V}_x$;
    else
      $z_v = 1$, $\forall\, v \in \mathcal{V}_x$;
      repeat
        $y_v = 1$, $\forall\, v \in \mathcal{V}_x$; $cond_{\text{h}} = 0$;
        repeat
          Compute $\tilde{U} : \Gamma(\boldsymbol{y}, \boldsymbol{z}, x, \tilde{U}) = 0$;
          $cond_{\text{m}} = 0$;
          for all $v \in \mathcal{V}_x : y_v z_v = 1$ **do**
            $\tilde{F}_v^{(x)} = \eta_v e^{\frac{\tilde{U}-1}{w_v a_v}} + b_v$;
            if $\tilde{F}_v^{(x)} < F_v^{\min}$ **then**
              $\tilde{F}_v^{(x)} = F_v^{\min}$; $y_v = 0$; $cond_{\text{m}} = 1$,
              $y_v = 0$;
              break
            end if
          end for
        until $cond_{\text{m}} = 0$
        for all $v \in \mathcal{V}_x : y_v z_v = 1$ **do**
          if $\tilde{F}_v^{(x)} > F_v^{\text{hl}}$ **then**
            $\tilde{F}_v^{(x)} = F_v^{\text{hl}}$; $z_v = 0$; $cond_{\text{h}} = 1$,$z_v = 0$;
          end if
        end for
      until $cond_{\text{h}} = 0$
    end if
    if $\sum_{v\in\mathcal{V}_x} w_v F_v^{(x)} \ge \sum_{v\in\mathcal{V}_{x+1}} w_v F_v^{(x+1)}$ **then**
      $x^* = x$;
    end if
  end if
  Transmit the video in the set $\mathcal{V}_{x^*}$ with rate $\tilde{F}_v^{(x^*)}$
end for

Algorithm 2 Pseudo code to solve problem (13)

Input: $\mathcal{V}$, R_0; $F_v^{\text{bl}}, F_v^{\text{hl}}, U_v^{\text{target}}, w_v, a_v, b_v, c_v, \forall v \in \mathcal{V}_N$;
for all $x = N, \dots, 1$ **do**
  if $\sum_{v\in\mathcal{V}_x} HF_v^{\min} \le R_0$ **then**
    if $\sum_{v\in\mathcal{V}_x} HF_v^{\text{hl}} \le R_0$ **then**
      set $\tilde{F}_v^{(x)} = F_v^{\text{hl}}$, $\forall v \in \mathcal{V}_x$;
    else
      $z_v = 1$, $\forall\, v \in \mathcal{V}_x$;
      repeat
        $y_v = 1$, $\forall\, v \in \mathcal{V}_x$; $cond_{\text{h}} = 0$;
        repeat
          $cond_{\text{m}} = 0$;
          for all $v \in \mathcal{V}_x : y_v z_v = 1$ **do**
            $\tilde{F}_v^{(x)} = w_v a_v \frac{\Lambda(\boldsymbol{y},\boldsymbol{z},x) - \sum_{n\in\mathcal{V}_x} y_n z_n b_n}{\sum_{n\in\mathcal{V}_x} w_n y_n z_n a_n} + b_v$
            if $\tilde{F}_v^{(x)} < F_v^{\min}$ **then**
              $\tilde{F}_v^{(x)} = F_v^{\min}$; $y_v = 0$; $cond_{\text{m}} = 1$;
              $y_v = 0$;
              break
            end if
          end for
        until $cond_{\text{m}} = 0$
        for all $v \in \mathcal{V}_x : y_v z_v = 1$ **do**
          if $\tilde{F}_v^{(x)} > F_v^{\text{hl}}$ **then**
            $\tilde{F}_v^{(x)} = F_v^{\text{hl}}$; $z_v = 0$; $cond_{\text{h}} = 1$,$z_v = 0$;;
          end if
        end for
        for all $i \in \mathcal{V}_x$ **do**
          for all $j \in \mathcal{V}_x : j > i$ **do**
            if $U_i(\tilde{F}_i^{(x)}) > U_j(\tilde{F}_j^{(x)}) \wedge w_i < w_j$ **then**
              $F_j^{\min} = \eta_v \exp(\frac{U(\tilde{F}_i^{(x)})}{a_v}) + b_v$;
              $cond_{\text{h}} = 1$;
            end if
          end for
        end for
      until $cond_{\text{h}} = 0$
    end if
    if $\sum_{v\in\mathcal{V}_x} w_v F_v^{(x)} \ge \sum_{v\in\mathcal{V}_{x+1}} w_v F_v^{(x+1)}$ **then**
      $x^* = x$;
    end if
  end if
  Transmit the video in the set $\mathcal{V}_{x^*}$ with rate $\tilde{F}_v^{(x^*)}$
end for

while a large value of α forces to select and transmit a lower number of videos with higher quality.

Finally note that constraint (3b) 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.

1) Solutions of problem (3) *and Algorithms:* By exploiting the invertible Q-R function in eq. (2), constraints (3b)-(3d) collapse in a unique constraint, *i.e.*,

$$F_v^{\min} < F_v < F_v^{\text{hl}} \qquad \forall v \in \mathcal{V}_x \tag{7}$$

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

$$F_v^{\text{target}} = \eta_v e^{\frac{U_v^{\text{target}}}{a_v}} + b_v \tag{8}$$

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 (3a), *i.e.*, the maximization with respect to the number x of video to transmit, is to exploit the feasibility condition resulting from the combination of the constraints (7) and (3c), *i.e.*,

$$\sum_{v\in\mathcal{V}_x} HF_v^{\min} \le R_0. \tag{9}$$

In fact, by recalling that $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots \subset \mathcal{V}_N$, the left-hand side of (9), *i.e.*, $\sum_{v\in\mathcal{V}_x} HF_v^{\min}$, is a strictly increasing sequence of the integer x. Hence, the condition (9) allows us to derive the maximum value of the integer x that makes the problem feasible. On the other hand, the minimum value of x can be derived from the trivial condition $\sum_{v\in\mathcal{V}_x} HF_v^{\text{hl}} > R_0$, *i.e.*, if x is too small and does not satisfy such condition, all the videos can be transmitted at the highest quality in real-time without adaptation.

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

$$\sum_{v\in\mathcal{V}_x} HF_v = R_0, \tag{10a}$$

$$\Delta_{\boldsymbol{p}}(U_i, U_j) = 0 \qquad \forall i, j \in \mathcal{V}, i > j \tag{10b}$$

$$F_v^{\min} < F_v < F_v^{\text{hl}} \qquad \forall v \in \mathcal{V} \tag{10c}$$

The optimal algorithm proposed in [33] to solve a problem similar to the inner maximization in (10) according to a MSE-based R-D model can be now suitably extended to consider the SSIM-dependent Q-R function, by replacing the functions of eq. (5) and (6) in [33] as follows:

$$\Gamma(\boldsymbol{y}, \boldsymbol{z}, x, U) = \sum_{v\in\mathcal{V}_x} \left[y_v z_v \left(\eta_v' e^{\frac{U-1}{w_v a_v}} + b_v \right) \right] - \Lambda(\boldsymbol{y}, \boldsymbol{z}, x) \tag{11}$$

where $\eta'_v = e^{\frac{1-c_v}{a_v}}$,

$$\Lambda(\boldsymbol{y}, \boldsymbol{z}, x) = \frac{R_0}{H} - \sum_{v \in \mathcal{V}_x} \left[(1 - y_v) F_v^{\min} + (1 - z_v) F_v^{\text{hl}} \right] \quad (12)$$

and $\boldsymbol{y} = [y_1, \ldots, y_N], \boldsymbol{z} = [z_1, \ldots, z_N]$ 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 be transmitted to search for the best video set $\mathcal{V}_{x^*}$ for which the weighted sum of the utility is maximum. We report the overall algorithm to solve the problem (3) in Algorithm 1, whose complexity can be estimated as follows. The inner loop (lines 8-32 of Algorithm 1) requires in the worst case $x(x+1)/2$ iterations, with $x \leq N$, whereas the outer loop requires in the worst cases N iterations. By neglecting the low-order terms, the worst-case complexity is $O(N^3)$. We refer the interested reader to [33] for further details.

B. Efficiency-Based Optimization

The proposed efficiency-oriented dynamic rate adaptation strategy consists in maximizing the weighted sum of the video utilities under the following constraints: available GBR, and maximum rate for each video, minimum utility and rate for each video. With respect to the optimization in (3) fairness is not taken into account, but a video utility ordering based on the visual quality is introduced. The novel problem is stated as follows:

$$\max_{1 \leq x \leq N} \max_{F_v, v \in \mathcal{V}_x} \sum_{v \in \mathcal{V}_x} w_v U_v(F_v) \quad (13a)$$

$$s.t. \sum_{v \in \mathcal{V}_x} H F_v \leq R_0, \quad (13b)$$

$$F_v^{\text{bl}} < F_v < F_v^{\text{hl}} \qquad \forall v \in \mathcal{V}_x \quad (13c)$$

$$(U_i - U_j)(w_i - w_j) \geq 0, \ \forall i, j \in \mathcal{V}_x \quad (13d)$$

$$U_v \geq U_v^{\text{target}}, \quad \forall v \in \mathcal{V}_x \quad (13e)$$

where the novel constraint (13d) 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 eq. (6).

1) Solutions of problem (13) *and Algorithms:* Since the objective is concave and the constraints convex, the inner maximization in (13) 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 (13d) would require the derivation of $x(x-1)/2$ dual variables, through, *e.g.*, sub-gradient method. In order to keep the complexity low, we consider here a low-complexity sub-optimal solution, which is derived by applying a method similar to the one proposed in the previous section, *i.e.*, by relaxing the reduced problem (13a)-(13b). In this case the solution without considering constraint (13d) is obtained by replacing line 14 of Algorithm 1 with the following equation:

$$\tilde{F}_v^{(x)} = w_v a_v \frac{\Lambda(\boldsymbol{y}, \boldsymbol{z}, x) - \sum_{n \in \mathcal{V}_x} y_n z_n b_n}{\sum_{n \in \mathcal{V}_x} w_n y_n z_n a_n} + b_v \quad (14)$$

Source and Coding Parameters	
Number of max. video per stage	5
Frame rate	30 fps
GOP Size	8
IDR period	32
EL encoding	MGS
Number of ELs	3
QP difference among ELs	4
System model	
User distribution	Uniform
Number of users	15
Number of best effort user	14
Cell layout	Single circular cell
Cell Range	300 m
Channel model	
Path Loss	$40 + 15.2 \log(d)$, d = distance in meter
Shadowing model	Log-normal with 4dB standard deviation
Channel Model	ITU A extended Pedestrian model
Doppler Bandwidth	5Hz
PHY model	
System Bandwidth	5 MHz
Subcarrier spacing	15 KHz
Number of subcarriers 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 II
SIMULATION PARAMETERS

Video Source	Position	Resolution	BL QP
Camera 1	outdoor	1024x720	46
Camera 2	outdoor	1024x720	46
Camera 3	outdoor	1024x720	46
Camera 4	outdoor	960x540	38
Camera 5	outdoor	720x574	38
Camera 6	indoor	640x480	46
Camera 7	indoor	640x480	46
Ultrasound	indoor	640x480	38

TABLE III
POSITION, RESOLUTION AND BL ENCODING QP OF THE CONSIDERED VIDEOS

as shown in Algorithm 2. An additional loop is considered to verify constraints (13d) with a resulting worst-case complexity of $O(N^4)$. Algorithm 2 lead to an optimal solution, if and only if

$$U_i(\tilde{F}_i^{(x)}) \geq U_j(\tilde{F}_j^{(x)}) \quad (15)$$

$\forall i, j > i \in \mathcal{V}_x : w_i > w_j, y_i z_i y_j z_j = 1$, meaning that the constraint (13d) is violated only for videos with minimum or maximum rate constraints active. In our numerical evaluation (15) always holds.

VI. NUMERICAL RESULTS

In our simulations we have considered an LTE-like single-cell access network with system bandwidth set to 5 MHz. Uplink single-carrier frequency division multiple access is assumed, where users may transmit over orthogonal physical resource block (PRB) with frequency spacing of 180 KHz and time duration of 1 ms. Radio resource allocation and scheduling are performed according to the methods proposed in [34], which consider both mixed and heterogeneous traffic.

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

.

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

TABLE IV
CRA RESULTS (q_v) FOR THE THREE CONSIDERED POSITIONS OF THE OBJECTS OF INTEREST

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[2] and receives from the access network a GBR R_0 with values negotiated in the range from 1 to 6 Mbps. All the other users are best-effort users, *i.e.*, $K_2 = K, K_1 = 0$. The radio channel for all users is modeled according to the ITU extended pedestrian A model, with a Doppler frequency of 5 Hz [40].

Since the focus of this work is on the combination of multistream video adaptation and multiuser resource allocation in emergency scenarios, we have assumed no packet-loss across the radio channel, and the resource allocation obtained by using the techniques and the algorithms in [21], [34] has been built on the channel capacity based rate estimation. In real LTE networks, near-zero packet-loss probability can be achieved by employing advanced adaptive modulation and coding techniques for Turbo codes, similar to what proposed in [41]. In addition, a packet-level FEC may be further applied at the higher layers of the protocol stack, thus enabling a complete packet-loss recovery at the cost of a slight reduction in the video throughput (see, *e.g.*, [42]).

[2]The ambulance is assumed to be statically located on the emergency area

We consider two different phases of the first-aid operations. As illustrated in Fig. 1, 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 Table III. Examples of frames acquired by the available devices are reported in Fig. 5.

Each video sequence is encoded with the JSVM reference software [43] with one BL and three SNR ELs. MGS is considered to encode the quality layers. Both enhancement frames and BL key-picture frames are used for motion compensation, thereby limiting the drift issue during the adaptation process [23]. As indicated in Table III, the quantization parameter (QP) of the ambient videos BL (indoor and outdoor) is set to 46, while the ultrasonography video BL is encoded with QP equal to 38. The QP difference between adjacent ELs 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 (2) 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 priority weight of the ultrasound video is set to 2. We consider a target video quality for the ultrasound video, 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 $H = 1$.

The proposed strategies for joint video selection and adaptation, denoted as prioritized equal quality (PEQ) (see section V-A) and maximum weighted quality (MWQ) (see Section V-B) 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 [26]. Moreover, ER does not perform context-aware video selection and prioritization.

In Table IV the visual quality metrics defined in Sec. III are reported for the outdoor ambient videos available in the first phase of the scenario. In particular, the values q_v 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 $\{4, 5, 1, 3, 2\}$. For the third object of interest only the first three cameras are useful, as the object is outside the FoV of *Camera 4* and *Camera 5*.

We investigate the impact of the shaping parameter α, used in the CRA-based camera weights as in eq. (6). In Fig. 6, the weighted sum of the utilities $\sum_{v\in\mathcal{V}_x} w_v U_v(F_v)$ (Fig. 6(a)), and the normalized sum of utilities $\frac{1}{x}\sum_{v\in\mathcal{V}_x} U_v(F_v)$ (Fig. 6(b)) are plotted with respect to the parameter α 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) 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 available 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 α is reasonable to be set up for general monitoring purposes, thereby ensuring the transmission of most of the ambient videos, but with low video quality. As the parameter α increases, the effect of the visual quality values 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. 6(b).

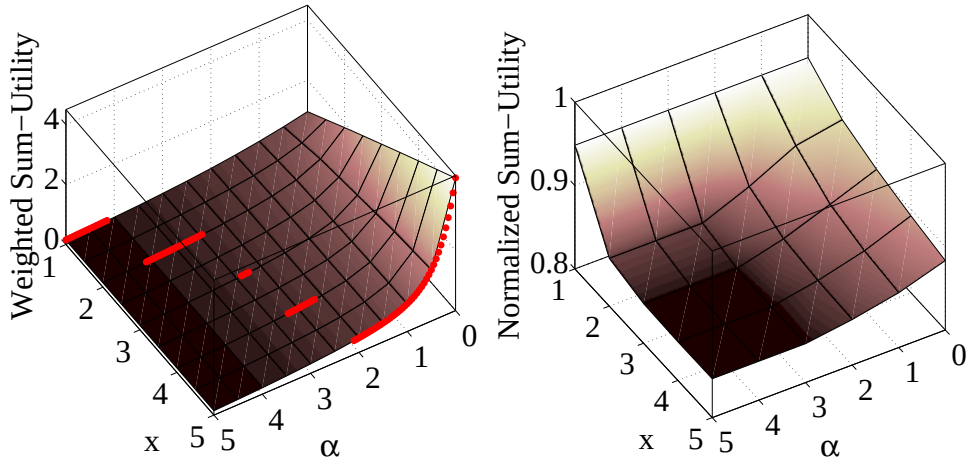
Fig. 7 shows the impact of the available GBR on the optimum number of video to transmit, *i.e.*, x^*, (Fig.7a) and on the normalized sum of the utility (Fig.7b), for different value of the shaping parameter α. We can note that the optimum number of video x^* does not significantly depend on the GBR, and remains approximately constant when α is fixed. Fig.7b highlights the benefit of decreasing the parameter α 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 we will use $\alpha = 2.5$.

In Fig. 8 we show the expected video quality, *i.e.*, the SSIM evaluated through the continuous model in (2), and the actual received video quality at the remote hospital, for all the ambient videos transmitted during the first phase. The three adaptation strategies proposed in the paper, namely, PEQ, MWQ and ER can be compared from the different subfigures. Three different objects of interest are considered, whose visual weights q_v have been reported in Table IV. In this case, the GBR $\bar{R}_0$ is set to 4 Mbps. We can note 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. 8a vs 8b) 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 improvements 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. 9 we report only the PEQ results. Even at large GBR, *i.e.*, $R_0 = 6$ Mbps (Fig. 9a) 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.*, $R_0 = 3$ 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.

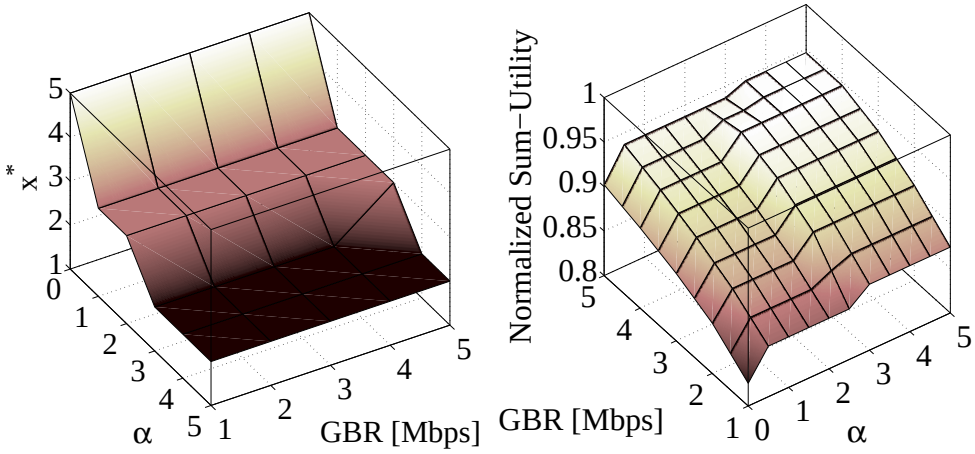
VII. Conclusion

We have presented a joint framework for context-aware selection and content-aware adaptation of a set of heterogeneous SVC videos transmitted over LTE uplink in a m-health emergency scenario. The proposed methods select the videos from one or more cameras taking into account a specific ranking criteria mainly related to the quality of the visual representation of the object of interest and adapt the video streams to the available bandwidth according to two different SVC-based adaptation strategies. The first quality-fair strategy aims at delivering the highest visual and video quality according to different priorities, derived from the CRA outcomes, and some fairness constraints. The second efficiency-based strategy aims at maximizing the aggregated weighted quality, where the weights are derived according to the visual ranking process. Both strategies are based on the modeling and evaluation of SSIM quality metric and have been tested and compared in the two operational phases of the considered emergency scenarios. Several numerical results are presented showing the significant improvements of the end-to-end video qualities of both diagnostic and ambient videos, compared to a benchmark adaptation strategy. Moreover, when the GBR granted to the ambulance decreases, the proposed strategies appropriately discard the videos coming from the



(a) (b)

Fig. 6. 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 α and the number of transmitted video x. The utilities are obtained from the inner maximization of problem (3) with $R_0 = 3$ Mbps. The markers in (a) indicate the optimal solutions for x.



(a) (b)

Fig. 7. 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 α and the GBR granted to the Ambulance.

cameras providing the lowest visual quality in order to improve the quality of the diagnostic videos, which are the most important for an effective tele-diagnosis.

REFERENCES

[1] G. Eysenbach, "What is E-Health?," *Journal of Medical Internet Research*, vol. 3, no. 2, pp. 20, Jun 2001.

[2] K. Perakis, *Third Generation (3G) Cellular Networks in Telemedicine: Technological Overview, Applications, and Limitations.*, IGI Global,, 2014.

[3] M.G. Martini, R. S H Istepanian, M. Mazzotti, and N.Y. Philip, "Robust multilayer control for enhanced wireless telemedical video streaming," *IEEE Trans. on Mobile Computing*, vol. 9, no. 1, pp. 5–16, Jan 2010.

[4] R. Paradiso, A. Alonso, D. Cianflone, A. Milsis, T. Vavouras, and C. Malliopoulos, "Remote health monitoring with wearable non-invasive mobile system: The healthwear project," in *IEEE 30th Int. Conf. on Engineering in Medicine and Biology Society, EMBS.*, Aug 2008, pp. 1699–1702.

[5] E. Kyriacou, M.S. Pattichis, C.S. Pattichis, A. Panayides, and A. Pitsillides, "M-health e-emergency systems: Current status and future directions [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 1, pp. 216–231, Feb 2007.

[6] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, pp. 1–22, 2009.

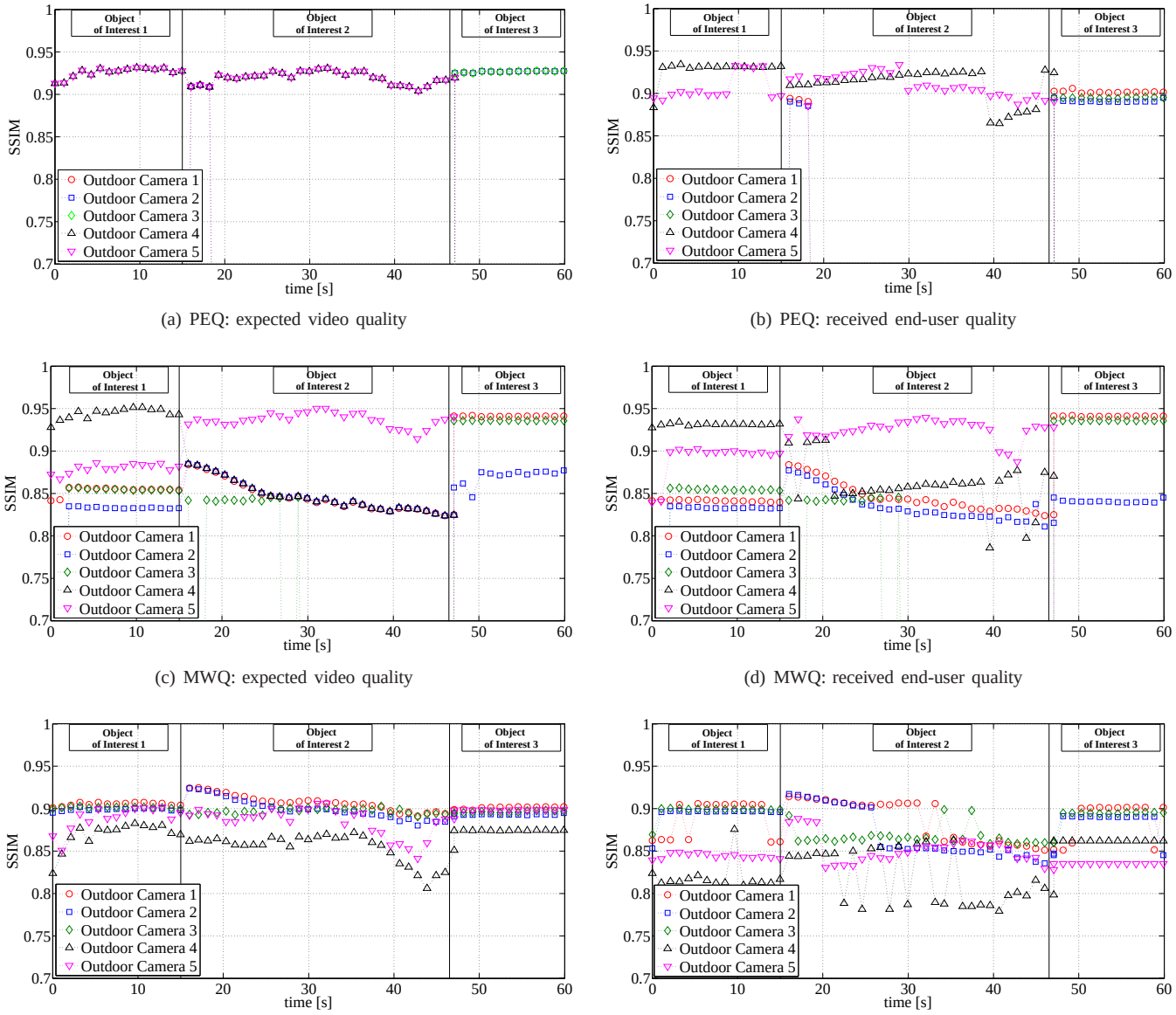
[7] R. Ferrus, O. Sallent, G. Baldini, and L. Goratti, "LTE: the technology driver for future public safety communications," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 154–161, Oct 2013.

[8] L. Skorin-Kapov and M. Matijasevic, "Analysis of QoS Requirements for E-Health Services and Mapping to Evolved Packet System QoS Classes," *Int. J. of Telemedicine and Applications*, vol. 9, 2010.

[9] S. Soro and W. Heinzelman, "Camera selection in visual sensor networks," in *IEEE Conf. on Advanced Video and Signal Based Surveillance, AVSS*, Sept 2007, pp. 81–86.

[10] K.R. Konda and N. Conci, "Optimal configuration of PTZ camera networks based on visual quality assessment and coverage maximization," in *7th Int. Conf. on Distributed Smart Cameras (ICDSC)*, Oct 2013, pp. 1–8.

[11] Y. Cho, S.O. Lim, and H.S. Yang, "Collaborative occupancy reasoning in visual sensor network for scalable smart video surveillance," *IEEE Trans. on Consumer Electronics*, vol. 56, no. 3, pp. 1997–2003, Aug



(a) PEQ: expected video quality

(b) PEQ: received end-user quality

(c) MWQ: expected video quality

(d) MWQ: received end-user quality

(e) ER: expected video quality

(f) ER: received end-user quality

Fig. 8. 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.

2010.

[12] J.R. 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 Trans. Inf. Technol. Biomed.*, vol. 9, no. 1, pp. 13–22, March 2005.

[13] A.S. Panayides, M.S. Pattichis, A.G. Constantinides, and C.S. Pattichis, "M-health medical video communication systems: An overview of design approaches and recent advances," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013*, July 2013, pp. 7253–7256.

[14] M.G. Martini, R.S.H. Istepanian, M. Mazzotti, and N. Philip, "A cross-layer approach for wireless medical video streaming in robotic teleultrasonography," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007*, Aug 2007, pp. 3082–3085.

[15] R.S.H. Istepanian, N.Y. Philip, and M.G. Martini, "Medical QoS provision based on reinforcement learning in ultrasound streaming over 3.5G wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 4, pp. 566–574, May 2009.

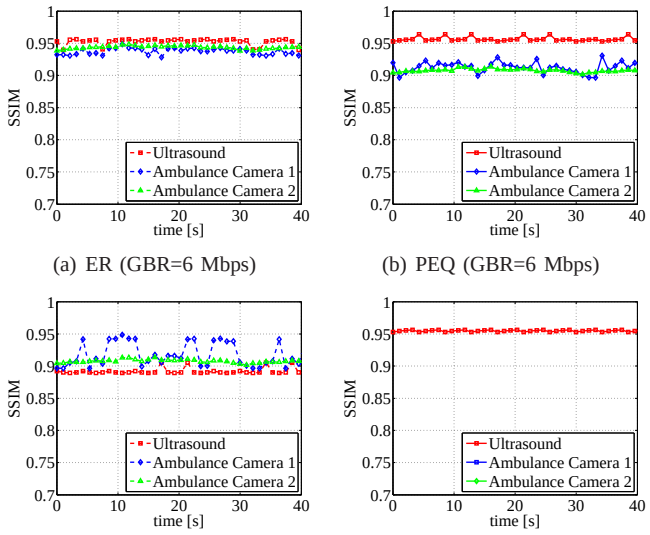
[16] G. Markarian, L. Mihaylova, D.V. Tsitserov, and A. Zvikhachevskaya, "Video distribution techniques over WiMAX networks for m-health applications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 24–30, Jan 2012.

[17] A. Alinejad, N.Y. Philip, and R.S.H. Istepanian, "Cross-layer ultrasound video streaming over mobile WiMAX and HSUPA networks," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 31–39, Jan 2012.

[18] I.U. Rehman and N.Y. Philip, "M-QoE driven context, content and network aware medical video streaming based on fuzzy logic system over 4G and beyond small cells," in *IEEE EUROCON 2015 - International Conference on Computer as a Tool*, Sept 2015, pp. 1–6.

[19] A. Panayides, Z.C. Antoniou, Y. Mylonas, M.S. Pattichis, A. Pitsillides, and C.S. Pattichis, "High-resolution, low-delay, and error-resilient medical ultrasound video communication using H.264/AVC over mobile WiMAX networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 619–628, May 2013.

[20] M.G. Martini, L. Iacobelli, C. Bergeron, C.T. Hewage, G. Panza, E. Piri, J. Vehkapera, P. Amon, M. Mazzotti, K. Savino, and L. Bokor, "Real-time multimedia communications in medical emergency - the CONCERTO project solution," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015*, Aug 2015, pp. 7324–



(a) ER (GBR=6 Mbps) (b) PEQ (GBR=6 Mbps)

(c) ER (GBR=3 Mbps) (d) PEQ (GBR=3 Mbps)

Fig. 9. 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

7327.

[21] L. Iacobelli, G. Panza, E. Piri, J. Vehkapera, M. Mazzotti, S. Moretti, S. Cicalò, L. Bokor, N. Varga, and M.G. Martini, "An architecture for m-health services: The CONCERTO project solution," in *European Conf. on Networks and Communications (EuCNC), 2015*, June 2015, pp. 118–122.

[22] Z. Antoniou, A.S. Panayides, M.S. Pattichis, S. Stavrou, E. Kyriacou, A. Spanias, A.G. Constantinides, and C.S. Pattichis, "Adaptive emergency scenery video communications using HEVC for responsive decision support in disaster incidents," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015*, Aug 2015, pp. 173–176.

[23] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[24] 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, pp. 428–397, 2008.

[25] H Hu, X. Zhu, Y. Wang, R. Pan, J. Zhu, and F. Bonomi, "Proxy-based multi-stream scalable video adaptation over wireless networks using subjective quality and rate models," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1638–1652, Nov 2013.

[26] S. Cicalò and V. Tralli, "Distortion-fair cross-layer resource allocation for scalable video transmission in OFDMA wireless networks," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 848–863, April 2014.

[27] 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.

[28] D.K. Shaeffer, "MEMS inertial sensors: A tutorial overview," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 100–109, April 2013.

[29] B. Sobhani, E. Paolini, A. Giorgetti, M. Mazzotti, and M. Chiani, "Target tracking for UWB multistatic radar sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 125–136, Feb 2014.

[30] D. Dardari, P. Closas, and P.M. Djuric, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1263–1278, April 2015.

[31] 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.

[32] 3GPP, "Policy and charging control architecture," TS 23.203, v10.7.0, 2012.

[33] 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 *2013 IEEE 15th International Conference on e-Health Networking, Applications Services (Healthcom)*, Oct 2013, pp. 212–217.

[34] S. Cicalò and V. Tralli, "Fair resource allocation with QoS support for the uplink of LTE systems," in *European Conf. on Networks and Commun. (EuCNC)*, June 2015, pp. 180–184.

[35] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[36] M. Razaak, M. G. Martini, and K. Savino, "A study on quality assessment for medical ultrasound video compressed via HEVC," *IEEE Journal on Biomedical Informatics*, vol. 18, no. 5, pp. 1552–1559, Sept 2014.

[37] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, 2000.

[38] S. Cicalò, N. Changuel, V. Tralli, B. Sayadi, F. Faucheux, and S. Kerboeuf, "Improving QoE and fairness in HTTP adaptive streaming over LTE network," *To appear in IEEE Trans. on Circuits and Systems for Video Technology*, 2015.

[39] A. Alinejad, N. Philip, and R.S.H. Istepanian, "Performance analysis of medical video streaming over mobile WiMAX," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2010, pp. 3471–3474.

[40] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); base station radio transmission and reception," TS 36.104 v10.1.0, 2010.

[41] M. Mazzotti, S. Moretti, and M. Chiani, "Multiuser resource allocation with adaptive modulation and LDPC coding for heterogeneous traffic in OFDMA downlink," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2915 –2925, october 2012.

[42] M. Mazzotti, E. Paolini, M. Chiani, B. Gadat, C. Bergeron, and R. Fracchia, "Analysis of packet-level forward error correction for video transmission," in *Proc. IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011*, May 2011, pp. 1–5.

[43] *JSVM 9.19.11 Reference Software Febrary 2011*.

Sergio Cicalò (S'10-M'14) received the M.Sc. Degree and the Ph.D. Degree from the University of Ferrara, Italy in 2010 and in 2014, respectively, all in Communication Engineering. From 2014 to 2016, He was a Researcher of the University of Ferrara. He is currently R&D Engineer at Elettronica Marittima S.r.l, Italy. He authored 15+ peer-reviewed publications in international Journal and Conference Proceedings and serves as a reviewer for IEEE Transactions/Journals and Conferences. His research interests are in the wide area of radar sensors, wireless communication and video delivery systems.

Matteo Mazzotti (M'08) received the Dr. Ing. degree (with honors) in telecommunications engineering and the Ph.D. in electronic engineering, computer science and telecommunications from the University of Bologna in 2002 and 2007, respectively. He has worked with the National Research Council (CNR), Italy, with the National Inter-University Consortium for Telecommunications (CNIT), Italy, and with the Center for Industrial Research on ICT (CIRI ICT) of the University of Bologna. He has participated in several national and international projects, such as the European projects Phoenix-FP6, Optimix-FP7, Concerto-FP7 and H2020 EuroCPS. Currently he is R&D Program Manager in OCEM Airfield Technology. He serves as a reviewer for international journals and conferences and has participated in the organizing committees and technical program committees of several international conferences. His main research interests include wireless multimedia communications, joint source and channel coding, broadcasting tecnologies, UWB radio localization systems, airfield ground lighting, monitoring and control technologies for AGL.

Simone Moretti was born in Rimini, Italy, on 29 June 1983. He received his Ph.D. in Electronics Engineering, Telecommunications and Information Technology in May 2015 from the University of Bologna, Italy. Currently, he is a Research Fellow at the University of Bologna, Italy. His main research interest cover multimedia data processing and communications, radio resource optimization, multi-source wireless communications systems, positioning and tracking.

Velio Tralli (S'93-M'94-SM'05) received the Dr.Ing. degree in electronic engineering (cum laude) and the Ph.D. degree in electronic engineering and computer science from the University of Bologna, Italy, in 1989 and 1993, respectively. From 1994 to 1999, he was a Researcher of the National Research Council (CNR) at CSITE, University of Bologna. In 1999, he joined the Engineering Department of the University of Ferrara, Italy where he is currently an Associate Professor. His research interests are within the areas of digital transmission and coding, and wireless communications, with emphasis on radio resource management, cross-layer design and optimization, OFDM and multiantenna systems, and cooperative communications. He participated in several national and European research projects addressing short-range communications systems, wireless sensor networks, 3G-4G wireless networks, wireless video communications. He published more than one hundred papers in refereed journals, including Transactions of the IEEE, and international conferences. He is a Senior Member of IEEE where he serves as a reviewer for Transactions/Journals and Conferences, and as a TPC member for several international conferences. He is an associate Editor for the European Transactions on Emerging Technologies. He also served as a Co-Chair for the Wireless Communication Symposium of ICC2006 and for the Communication Theory Symposium of ICC2013.

Marco Chiani (M'94SM'02F'11) received the Dr. Ing. degree (summa cum laude) in electronic engineering and the Ph.D. degree in electronic and computer engineering from the University of Bologna, Italy, in 1989 and 1993, respectively. He is a Full Professor in Telecommunications at the University of Bologna. During summer 2001, he was a Visiting Scientist at AT&T Research Laboratories, Middletown, NJ. Since 2003 he has been a frequent visitor at the Massachusetts Institute of Technology (MIT), Cambridge, where he presently holds a Research Affiliate appointment. He is leading the research unit of the University of Bologna on cognitive radio and UWB (European project EUWB), on Joint Source and Channel Coding for wireless video (European projects Phoenix-FP6, Optimix-FP7, Concerto-FP7), and is a consultant to the European Space Agency (ESA-ESOC) for the design and evaluation of error correcting codes based on LDPCC for space CCSDS applications. His research interests are in the areas of wireless systems and communications theory, including MIMO statistical analysis, codes on graphs, wireless multimedia, cognitive radio techniques, and ultra-wideband radios. He received the 2011 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems, the 2012 IEEE Communications Society Fred W. Ellersick Prize, and the 2012 IEEE Communications Society Stephen O. Rice Prize in the Field of Communications Theory. He is the past chair (20022004) of the Radio Communications Committee of the IEEE Communication Society and past Editor of Wireless Communication (20002007) for the journal IEEE Transactions on Communications. In 2012 he has been appointed Distinguished Visiting Fellow of the Royal Academy of Engineering, UK. Since 2011 he is a Fellow of the IEEE, named for Contributions to wireless communication systems".