A real-time video quality estimator for emerging wireless multimedia systems

Elisangela Aguiar · André Riker · Eduardo Cerqueira · Antônio Abelém ·
Mu Mu · Torsten Braun · Marilia Curado · Sherali Zeadally

Abstract Wireless Mesh Networks (WMNs) are increasingly deployed to enable thousands of users to share, create, and access live video streaming with different characteristics and content, such as video surveillance and football matches. In this context, there is a need for new mechanisms for assessing the quality level of videos because operators are seeking to control their delivery process and optimize their network resources, while increasing the user's satisfaction. However, the development of in-service and non-intrusive Quality of Experience assessment schemes for real-time Internet videos with different complexity and motion levels, Group of Picture lengths, and characteristics, remains a significant challenge. To address this issue, this article proposes a non-intrusive parametric real-time video quality estimator, called MultiQoE that correlates wireless networks' impairments, videos' characteristics, and users' perception into a predicted Mean Opinion Score. An instance of MultiQoE was implemented in WMNs and performance evaluation results demonstrate the efficiency and accuracy of MultiQoE in predicting the user's perception of live video streaming services when compared to subjective, objective, and well-known parametric solutions.

Keywords Video streaming · Quality of experience · Video quality estimator · Wireless mesh networks

1 Introduction

The last few years have witnessed a phenomenal growth in the wireless industry, both in terms of multimedia mobile technology and its human-centric subscribers. The current trends and demands in wireless communications require the

E. Aguiar · E. Cerqueira (✉) · A. Abelém
Faculty of Computer Engineering, Federal University of Para,
Belém, BR, Brazil
e-mail: cerqueira@ufpa.br

E. Aguiar
e-mail: eaguiar@ufpa.br

A. Abelém
e-mail: abelem@ufpa.br

A. Riker · M. Curado
Department of Informatics Engineering, University of Coimbra,
Coimbra, Portugal
e-mail: ariker@dei.uc.pt

M. Curado
e-mail: marilia@dei.uc.pt

E. Cerqueira
Computer Science Department, University of California Los
Angeles, Los Angeles, CA, USA

M. Mu
School of Computing and Communications, Lancaster
University, Lancaster, UK
e-mail: m.mu@lancaster.ac.uk

T. Braun
Institute of Computer Science and Applied Mathematics,
University of Bern, Bern, Switzerland
e-mail: braun@iam.unibe.ch

S. Zeadally
College of Communication and Information, University
of Kentucky, Lexington, KY 40506, USA
e-mail: szeadally@uky.edu

delivery of real-time video applications over heterogeneous wireless networks with Quality of Experience (QoE) [1, 2] support. Video streaming will provide new sources of income for network operators and content providers, since it will be a major application in future wireless systems and a key factor in ensuring their success [3, 4]. At the same time, users have been producing, sharing, and accessing thousands of video services on wireless devices.

Real-time multimedia traffic consists of one or more media streams with different spatial and temporal (motion and complexity) video activities and features. A Group of Picture (GoP) is a group of successive pictures within a coded video stream and composed of a combination of three frame types for the compressed video streams, namely I (Intra-coded), P (Predictive-coded), and B (Bidirectionally predictive-coded) frames. It is important to highlight that not all video frames are equal or have the same degree of importance from the user's point-of-view. For instance, depending on the video motion and complexity levels (e.g., a small moving region of interest on a static background or fast-moving sports clips) and the GoP length, the impact of a packet lost in the Human Visual System (HVS) may or may not be annoying [5].

Different types of wireless technologies, such as Wireless Mesh Networks (WMNs) [6], can be used to deliver a wide range of real-time video streaming services to a large number of users. However, video streaming produces a degraded performance in wireless systems, including in WMNs, due to network/channel impairments, such as packet loss [7, 8]. Understanding and modeling the relationship of network impairments, video characteristics, and human experiences by using wireless quality level assessment schemes are key requirements for the delivery of visual content in multimedia mobile networks, such as football matches and other live multimedia events [9]. The operators that assess the QoE of real-time video services have a significant advantage by being able to strike an ideal balance between network provisioning, video codec configuration, and user's experience.

Solutions for assessing the QoE of a video service can be organized as subjective or objective [10], where the latter is hard to implement in real-time. Objective video media quality assessment technologies are categorized into several parametric model types [9], where packet-layer schemes have been gaining attention due to their high accuracy and low processing. Packet-layer models predict the perceived video quality level based on information about the IP and video codec headers, such as frame type and packet loss rate (without decoding the video flow). The impact on user perception of video flows is influenced by the number of the edges (spatial information—complexity level) in the individual frames and by the type and direction of movement (temporal information—motion level) in a GoP. However,

existing solutions have not been implemented and evaluated in multimedia wireless systems, or presented inaccurate results from the user's experience, because, mainly, they do not consider the video motion and complexity levels in their assessment procedures.

This article extends our previous work [11] with a modular and parametric packet-layer wireless video quality estimator, called MultiQoE. MultiQoE uses IP and MPEG packet header information to predict the quality level of a variety of videos (different genres and content), which reduces the system complexity and processing. Without decoding the videos, MultiQoE estimates the quality of level live video sequences by orchestrating and mapping information on networks' impairments, videos' characteristics, and users' perception into a predicted Mean Opinion Score (MOS). In contrast to existing works, MultiQoE also uses information on motion and complexity levels of the video frames in a GoP to improve the system accuracy. Simulation experiments with the assistance of real viewers were carried out to demonstrate the benefits and evaluate the efficiency of MultiQoE in a wireless mesh multimedia network. The results show that MultiQoE predicts the quality level of a set of videos closest to the human experience when compared to other widely-used QoE video quality estimator models, such as the Peak Signal-to-Noise Ratio (PSNR) [12], Video Quality Metric (VQM) [13], Structural Similarity Index Metric (SSIM) [14], and Pseudo-Subjective Quality Assessment (PSQA) [15].

The remainder of the article is structured as follows. Section 2 describes the related works. The MultiQoE proposal is explained in Sect. 3. Section 4 presents the test environment, scenario, implementations, and simulation results. Some concluding remarks are summarized in Sect. 5.

2 Related works

The ITU-R Recommendation BT.500 [16] has defined subjective assessment as the most reliable system of Video Quality Assessment (VQA). Subjective methods measure the overall perceived video quality, under well-defined and controlled conditions, by asking observers to evaluate videos [17]. However, subjective assessment is cumbersome, expensive, and unsuitable for in-service and real-time applications [18, 19]. The most widely-used subjective scheme for video quality evaluation is MOS [16, 20] which is recommended by the ITU-Telecommunication (ITU-T) Standardization Sector. The MOS rates the video quality on a scale of 1–5, where 5 is the best possible score. It should be noted that in the tests, observers tend to avoid scores at the extreme end of the scale (1 or 5) due to the influence of psychological factors [21]. However, MOS is not suitable for real-time video assessment approaches.

In ITU-T Study Group (SG) 12, there is a study [22] on the non-intrusive objective parametric and well-structured QoE assessment models (e.g., G.OMVAS [22], P.NAMS [23] and P.NBAMS [24] as planning, packet-layer and bit stream models, respectively) that can predict the perceptual impact of network impairments on video applications, considering the kind of impairment caused by both transmission and video compression issues [7, 25]. The prediction is based on packet header information [26, 27] and prior knowledge of the media stream [28]. However, in practice, existing solutions [22–27] have not been implemented and validated in wireless multimedia systems, where the mapping of packet/network information into MOS is required. MultiQoE follows the ITU-SG 12 recommendations, defines its specific input video/packet/network parameters, and validates an accuracy parametric video quality estimator solution for multimedia WMNs.

The most popular objective quality inference techniques include PSNR [12], VQM [13], and SSIM [14]. Although attempts to assess coding quality have often focused on estimating the PSNR, the PSNR by itself, does not always correlate well with perceived quality of the HVS [25, 28, 29]. PSNR can only be computed once the image is received, which is not appropriate for real-time prediction systems [29–31].

VQM provides a better indication of perceptual quality than PSNR [32]. In general, VQM's ability to loosely classify a sequence as 'very poor' or 'very good' is accurate, but it often fails to distinguish sequences that share similar levels of degradation. The VQM evaluation results vary from 0 to 5, where 0 is the best possible score. Another well-known metric is called SSIM, which is based on a frame-to-frame measurement of three components (luminance, contrast, and structural similarity). The SSIM index is a decimal value between 0 and 1, where 0 means there is no correlation with the original image, and 1 means it has exactly the same image. However, both VQM and SSIM metrics cannot be used in real-time and perform poorly compared to MOS.

A non-intrusive QoE parametric scheme, called PSQA, has been used to predict the quality level of videos flows [15]. The authors included a Random Neural Network (RNN) model (together with its learning algorithms) to assess the quality level of videos in real-time based on a set of parameters, including frame type, frame rate, and packet loss rate. The proposed solution was originally proposed to improve the understanding of Quality of Service (QoS) factors in multimedia engineering without an in-depth understanding of actual user experience (lack of QoE support).

To assess the QoE of Multiple Description Coding (MDC) videos over multiple overlay paths, the proposal [33] was compared to subjective (e.g., MOS), objective (e.g., PSNR), and non-intrusive parametric approaches (e.g., PSQA). MDC-PSQA extends PSQA, together with QoE prediction, by also using the GoP length and the percentage losses of the I, P, and B frames in a GoP. However, it only considers one video (Foreman—no motion and complexity variation), which makes the system less accurate in assessing the quality level of videos with different characteristics, such as those expected for the Internet. Furthermore, visual quality metrics must be tested on a wide variety of visual contents and distortion types before meaningful conclusions can be drawn from their performance.

As presented in [5], the impact of the video quality level on HVS is highly influenced by the number of the edges (spatial information—complexity level) in the individual frames and by the type and direction of movement (temporal information—motion level). However, in contrast to MultiQoE, the PSQA mechanism and its extensions do not consider a set of diverse videos (they use only one video) with different levels of spatial temporal activities during their training and prediction procedures which reduce the system accuracy for measuring the quality level of (many) Internet videos.

Another proposal investigates the dependence of video quality on numbers, expressed as MOS for a given set of QoS network parameters [34]. This work investigates the impact of key frames on the quality perceived by users in wireless systems. Unlike MultiQoE, it only considers one video flow (and not videos with different motion and complexity levels) and does not take the GoP length into account, which is an important input in a QoE evaluation system.

The solution proposed by Khan et al. [35] has classified videos into groups representing different content types (using a combination of temporal and spatial levels) and extraction features, by means of cluster analysis. Based on the content type, the video quality (in terms of MOS) was predicted from the network parameters (e.g., packet error rate) and application-level parameters (e.g., transmission bit rate and frame rate) by using Principal Component Analysis (PCA). In contrast to MultiQoE, the proposed solution measures the video quality level by applying the average PSNR to all the decoded frames which performs poorly compared to MOS. Other extensions of this work [3, 36, 37] failed to provide satisfactory MOS results because PSNR was still used to correlate the video's characteristics and network's impairments into human scores.

Few works have analyzed the impact of the distribution of videos with different motion and complexity levels over wireless networks according to the human's perception (e.g., 38, 39). It is clear that the accuracy and performance of a non-intrusive parametric QoE video quality estimator is largely dependent on the video characteristics, including
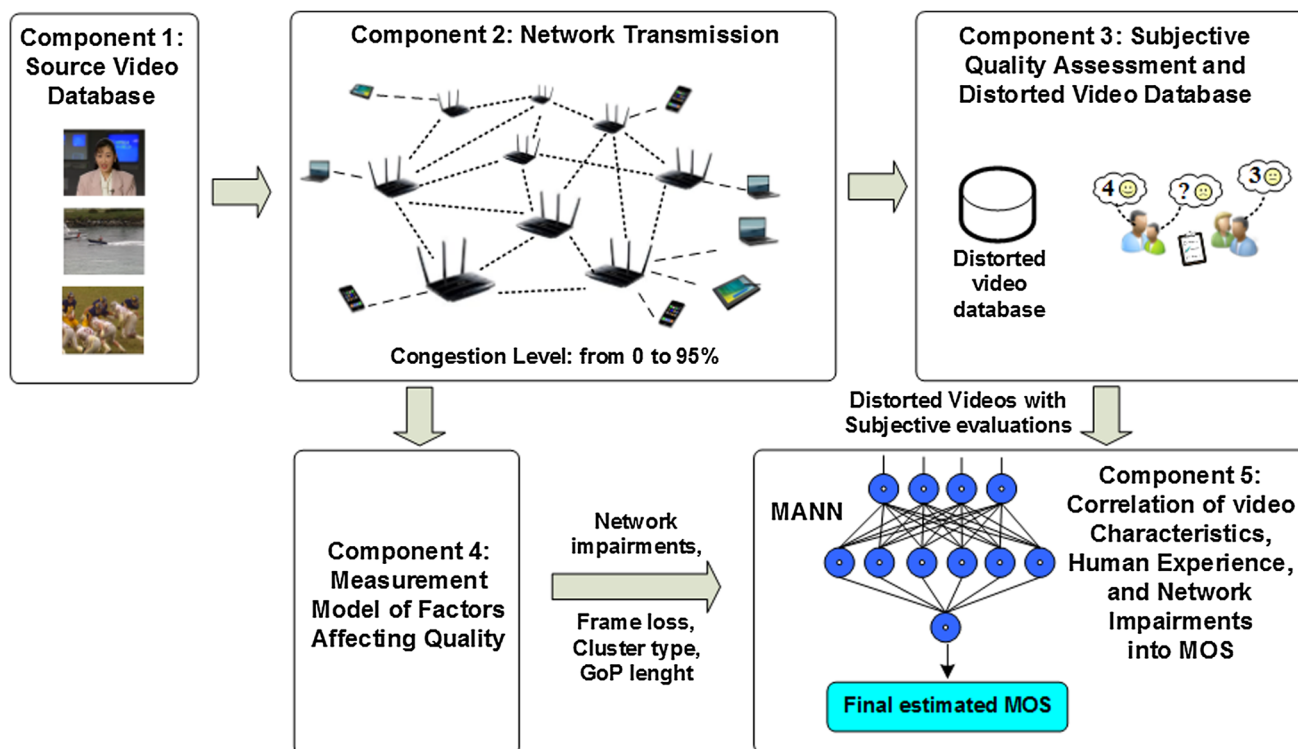
Fig. 1 MultiQoE components (see in annex as recommended by the journal)

GoP length, frame type in a GoP, motion and complexity vectors. A QoE video quality estimator also requires a good mapping technique to correlate the video content features and wireless network's impairments in the human scores, such as a cluster-based Artificial Neural Network (ANN) approach [40].

ANN has the ability to learn complex data structures and approximate a given continuous mapping. ANNs work fast (after a training phase) even with large amounts of data [41] and approximate a continuous mapping to any arbitrary degree of accuracy as expected for QoE-aware video prediction schemes. This means that they are suited to learning the salient characteristics of human perception [41] during the video quality estimation process. Existing image quality assessment schemes have demonstrated the benefits and feasibility of ANN-based models in predicting the quality level of images (not videos) [41–46].

Many existing works only use simple objective QoE metric to validate their proposals or fail to take into account the fact that videos have different levels of motion and complexity. Another advantage of MultiQoE is the use of a Multiple ANN (MANN) correlation approach to map video's characteristics, human's perceptions, and network's impairments into a predicted MOS with results close to human scores. MANN has been successfully used for QoS/QoE assessment schemes [3, 47, 48] and yielded better results than RNN and other techniques. MultiQoE

has a realistic assumption that not all packets are equal or have the same degree of importance which are key parameters to determine the extension of the video impairment, as discussed in [5, 49] (see Sect. 3.5).

3 MultiQoE

MultiQoE is a modular and flexible in-service parametric approach to predict the quality level of different video sequences, where it can be configured with any wired or wireless technologies with low complexity and high accuracy. However, due to the popularity of WMNs, the remainder of this article will present the use of MultiQoE in a multimedia WMN environment.

An instance of MultiQoE can be obtained by following the procedures defined in five main components. Each of them is designed to complete single or multiple tasks for the modeling of the quality evaluation model. The components of MultiQoE are illustrated in Fig. 1 and are as follows: (i) Source Video Database, (ii) Network Transmission; (iii) Subjective Quality Assessment and Distorted Video Database; (iv) Measurement Model of Factors Affecting Quality; and (v) Correlation of Video Characteristics, Human Experience, and Network Impairments into MOS.

The Source Video Database (Component 1) classifies typical Internet videos according to their spatial and
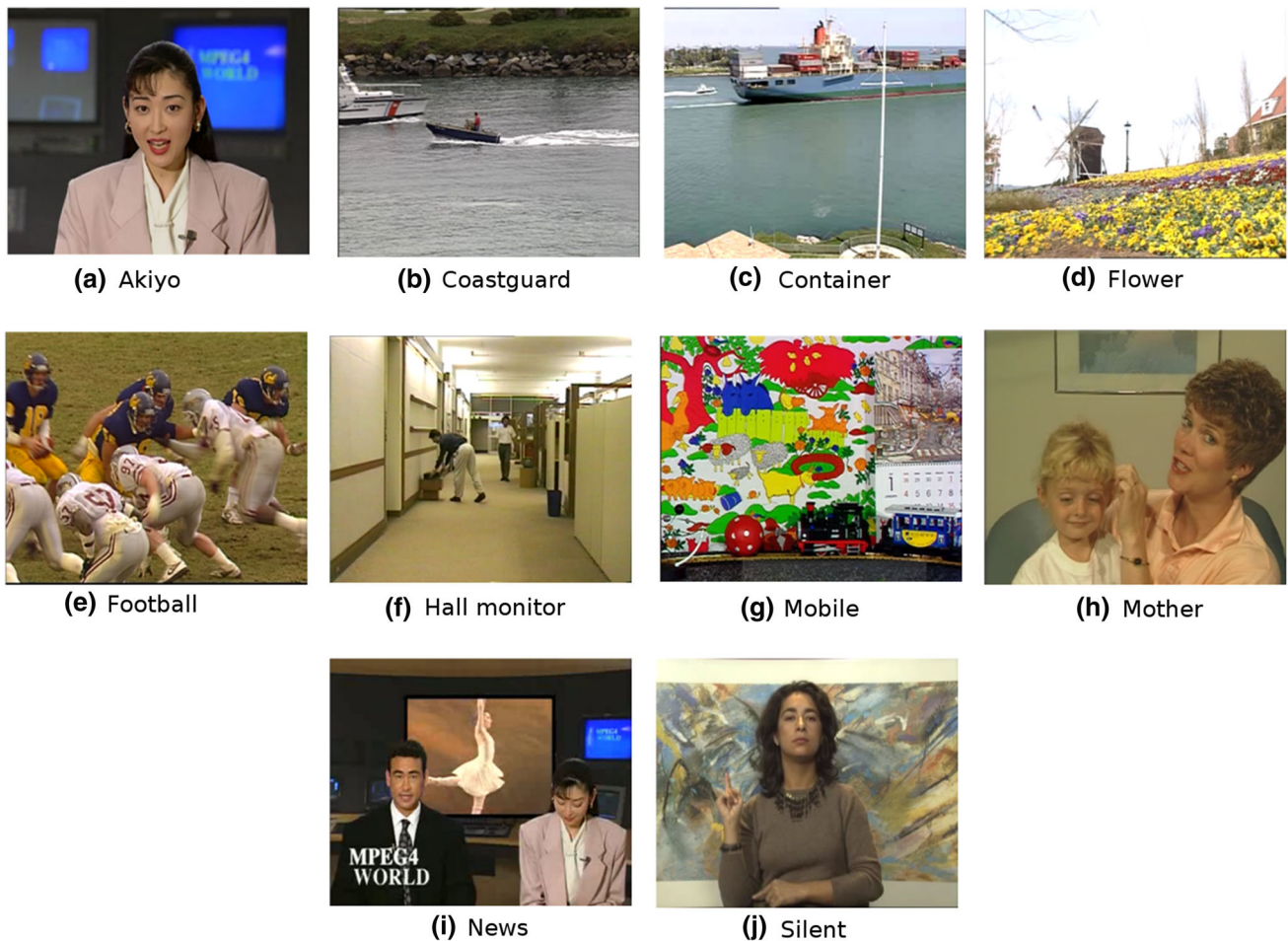
Fig. 2 Snapshots of the selected videos (see in annex as recommended by the journal)

temporal characteristics. The video content characteristics taken together with the percentage of losses of *I*, *P*, and *B* frames of a certain GoP (to improve the system accuracy, each ANN is responsible for videos with a specific GoP length, such as 10, 20, or 30) are used by Component 4 (Measurement Model of Factors Affecting Quality) to identify the video motion and complexity levels as well as the impact of the transmission on the video frames. At the same time, it is important to keep a distorted video database composed of videos delivered (as expected to be received/viewed by humans) in real/simulated networks (such as WMNs).

The Component 2 (Network Transmission) is responsible for transmitting all videos in wireless networks (with different congestion levels, errors, and impairments), getting information on packet loss and delay of video frames. The output of this component is important to create a distorted video database, where the videos experienced different network impairments, as well as measure the percentage of losses of the *I*, *P*, and *B* frames in a GoP as specified in Component 4. In the Component 3 (Subjective

Quality Assessment and Distorted Video Database), a panel of humans evaluates all distorted videos (following the ITU recommendations) to define/score their MOS. Finally, Component 5 uses a MANN to correlate video's characteristics, human' experience, and network's impairments into a predicted MOS.

3.1 Component 1: source video database

This component is responsible for maintaining videos with different features in a video database using a content classifier to classify the spatial and temporal characteristics of the videos. Information about spatial (edges and colors) and temporal (movement speed and direction) activity has been widely recognized as a key metric that can be used as input for VQA [50].

MultiQoE uses uncompressed sequences of natural scenes that are available in [51] to set up a video source database, containing video scenes of different characteristics ranging from very small movements (e.g., a small moving region of interest on a static background) to fast-

Table 1 DCT coefficients and MV of the selected videos

| Video	I Frames	P Frames		B Frames	
	DCT coefficients	DCT coefficients	MV	DCT coefficients	MV
Akiyo	80,104	9,213	488.5	1,340	308.1
Container	80,977	8,627	759.7	1,598	452.1
Hall	73,084	8,512	716.8	2,071	386.7
Mother	67,100	8,287	1,131.6	988	703.4
News	70,802	9,792	893.7	1,894	683
Silent	76,399	8,429	1,055.7	1,364	911.2
Mobile	88,911	22,310	1,473.2	7,583	971.1
Flower	82,163	24,479	1,374.2	7,325	998.9
Coastguard	39,682	26,302	1,969.8	8,600	3,091.8
Football	42,394	14,213	1,743.7	2,820	877.9

moving sports clips. All flows have realistic streaming sequences, with different types (e.g., news and football matches) and levels of complexity and motion, representing typical examples of content that are either distributed by providers, created/shared by end-users, and made available on the Internet.

Figure 2 shows one frame from each video used for this instance of MultiQoE. The videos are commonly used by many related works [52] and by the Video Quality Experts Group (VQEG) in many QoE experiments. Other videos with different resolution, durations, and characteristics can be easily included in MultiQoE. For instance, Component 1 could be composed of real user-generated Youtube or content-generator media videos. The network administrator can configure MultiQoE in accordance with his/her video features, network technologies, and business models.

For this proposal, MultiQoE uses 10 representative MPEG-4 digital videos, in YUV 4:2:0 format with a duration varying from 10 to 12 s to avoid the forgiveness effect. Owing to a restricted bandwidth, the test sequences only contain video flows and all of them were displayed in a native resolution of 352×288 pixels and 25 frames per second (fps).

The MPEG standard [53] defines three frame types for the compressed video streams, namely I, P, and B frames. The successive frames between two succeeding I frames define a GoP. A GoP pattern is characterized by two parameters as follows: GoP (N, M), where N is the I to I frame distance and M is the I to P frame distance. The encoding/decoding correlation between the frames, in particular, the B and P frames, depends on the respective preceding and succeeding I or P frames.

With a given video codec, the GoP structure and length can be configured according to the test plans. For this use case, the internal GoP structure was fixed by using two B frames for each P frame, as is the case in typical video streaming (IBBP pattern). Notice that there is no fixed
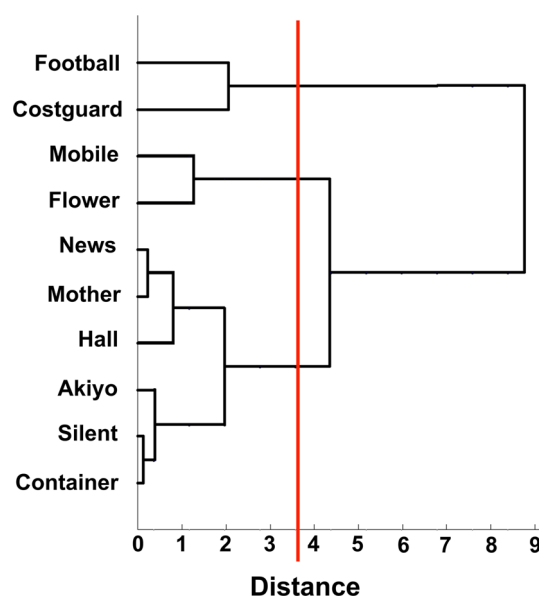
Fig. 3 Tree diagram based on a cluster analysis (see in annex as recommended by the journal)

default value for the GoP length, but the typical minimum and maximum values for resolution industrial video, such as MPEG-2 videos, are between 10 and 20. Therefore, we configured GoP lengths of 10, 20, and 30 in our experiments.

MultiQoE uses a content classifier as a function to classify the spatial and temporal characteristics of the videos as presented in Table 1. It is important to extract the Discrete Cosine Transform (DCT) coefficients and the Motion Vector (MV) sizes of the videos [54] without decoding the video payload. The content classifier carries out an evaluation of content-aware quality, which differentiates between the influence of video content characteristics and the perception of video quality. For instance, the

Table 2 Characteristics and classification of the clusters

Video	Video features	Cluster type
Akiyo, container, hall, mother, news and silent	Small moving region of interest on a static background	Low spatial temporal (LST)
Mobile and flower	Contiguous camera and motion; scenes with many small details	High spatial medium temporal (HSMT)
Coastguard and football	Fast camera motion or a rapidly shifting background	Medium spatial high temporal (MSHT)

video called *Mobile* has the largest DCT coefficients, and thus, the highest spatial complexity, while the video *Football* has the largest motion vector and the highest temporal activity.

MultiQoE uses a hierarchical clustering system based on nearest-neighbor Euclidean distance to classify the video content with similar spatial and temporal levels (i.e., according to the DCT and MV sizes). Figure 3 illustrates the cluster/multilevel hierarchy that was obtained after all the video transmissions, and shows how the videos were grouped in levels of proximity based on the DCT coefficient and the MV sizes. In accordance with the linkage distance, each video in the hierarchical tree is linked to the video or group that is most similar to it. MultiQoE uses three of the largest linkage distances to determine the cluster divisions in the data set (indicated by the red line in Fig. 3) and produce three clusters with high values of content similarity.

In this article, the videos were divided into three groups with the number of clusters being adjusted to different scenarios or application requirements. The number of clusters is specified so that a balance can be maintained between the effectiveness and complexity of the modeling process. The name of the clusters gives an indication of the aspects of the content in the videos that are most prevalent; however, an exhaustive classification is beyond the scope of this article. Furthermore, the number of clusters created can vary in accordance with the objectives of the system.

On the basis of what is shown in Fig. 3, and an analysis of the DCT coefficients and MV sizes, it is possible to classify the three clusters (see Table 2) as Low Spatial Temporal (LST), High Spatial Medium Temporal (HSMT), and Medium Spatial High Temporal (MSHT). Videos with small regions of interest and a static background compose the LST cluster. The HSMT cluster has videos where the camera is in constant motion and the scenes provide much more visual information. Finally, the MSHT cluster contains videos with fast camera or background motions and has scenes with an average amount of visual information.

**Fig. 4** UFPA mesh backbone (see in annex as recommended by the journal)

3.2 Component 2: network transmission

The Component 2 generates congestions/errors/impairments in multimedia networking environments, by using a network simulator (could be a testbed or network emulator). Thus, it will be possible to understand and model the relationship between network's impairments and user experience on the delivered/received video flows. The Component 3 uses the output of this component to maintain a database with videos transmitted over wireless links with different congestion rates and errors, as expected in real systems. Each received video is linked to a table with information about the losses of its *I*, *P*, and *B* frames.

To validate Component 2, WMN scenarios were simulated on the basis of the topology of a real testbed located at the Federal University of Para (UFPA) WMN backbone, in Amazon/Brazil, as shown in Fig. 4.

UFPA has dozen buildings distributed on its main campus. The physical characteristics of the campus, such as the number of trees and riverside setting, added to the fact that city of Belem/Amazon area experiences the high atmospheric humidity and frequent and often intense rainfall, make it a more challenging scenario than some of those described in the literature [55] for studying mesh networks. To represent a real scenario, a wireless device that was placed at random in the network simulation set up and received real videos transmissions from one source.
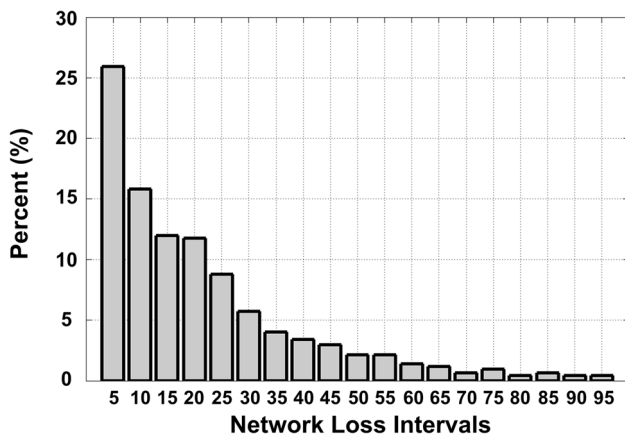
Fig. 5 Percentage of losses for each network interval (see in annex as recommended by the journal)

Each video streaming transmission starts by following a Poisson distribution.

The Gilbert-Elliot model, based on a two-state Markov chain, was used to generate loss in the WMN scenarios. Although important, the focus of the article is not on the transmission channel model. Congestion levels of up to 95 % were applied in the network (as shown in Fig. 5), where most of the videos experienced a network loss of up to 10 %, while a few had a loss of more than 50 %. Moreover, 900 experiments were carried out in the above scenarios to investigate the impact of different network's impairments on the quality level of different videos. Thus, it is possible to have a large and heterogeneous distorted video database, with videos transmitted over wireless links with different rates of congestion and errors.

It is well known that the loss propagation, as well as the impact on the user's experience caused by a packet dropped/error depend on the type of the frame in which the loss occurs, and on the GoP structure (as explained in Sect. 3.1). If an I or P frame is affect by loss, the loss propagates until the next I frame. If a B frame is affected, the loss does not propagate, except if it serves as a reference-point in the context of hierarchical coding [27, 56]. Thus, MultiQoE uses the percentage of losses of the I, P, and B frames, the GoP structure, and video content characteristics as input for the video quality estimator.

3.3 Component 3: subjective quality assessment and distorted video database

Upon generating a video distorted database, an evaluation scheme had been carefully implemented to assess the quality level of the videos by conducting subjective experiments. The experiments were carried out by asking a panel of observers to classify the quality level of distorted

videos by means of the test method laid down by the relevant ITU recommendations [57]. There are several recommendations [16–18] that stipulate strict conditions that have to be complied with to carry out subjective tests.

For the subjective test evaluation, MultiQoE uses the Single Stimulus (SS) method of ITU-R BT.500 [16] and a sample of 55 observers to collect the MOS results. They had normal vision and their ages ranged from 18 to 45. In the SS studies, the videos are only shown to the observers one at a time. MultiQoE uses a SS paradigm because it is ideally suited to a large number of video experiments [58]. Additionally, it significantly reduces the amount of time needed to conduct the study (given that there is a fixed number of subjects), compared to a Double Stimulus (DS) study [17].

Within a voting time of up to 5 s, observers assess the video quality level by selecting a score in the range of 1–5 which is combined with the quality scale (*Bad*; *Poor*; *Fair*; *Good* and *Excellent*). The study comprised a set of sequences shown in random order for each observer, as well as for each session. The Absolute Category Rating (ACR) method collects the opinions of the observers [16], because it provides a better replication of the streaming scenario in the real world [58] and is suitable for large-scale experiments that involve a large number of video flows.

For each video in the database, the mean score, $\bar{u}_{jkr}$, was calculated by:

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^{N} u_{ijkr}$$

where u_{ijkr}: score of observer i for test condition j, video sequence, k, repetition r. N: number of observers.

In the next stage, all the mean scores obtained are combined with a confidence interval, which is derived from the standard deviation and the size of each video. MultiQoE uses a 95 % confidence interval which is given by

$$\left[\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} - \delta_{ijkr}, \right]$$

where

$$\delta_{jkr}, = 1,96 \frac{S_{jkr}}{\sqrt{N}}$$

The standard deviation for each video, S_{jkr}, is given by

$$S_{jkr} = \sqrt{\sum_{i=1}^{N} \frac{\bar{u}_{jkr} - u_{jkr}}{(N-1)}}$$

After calculating all the scores, the video database is prepared to program (be an input) the MANN. Figure 6 shows a histogram that displays the obtained MOS scores. The score 3 (*fair*) is the value that is most often selected by the evaluators (35.79 %), while the grade 2 (*poor*), 30.21 %, and 4 (*good*), 16.53 % are the second choice and
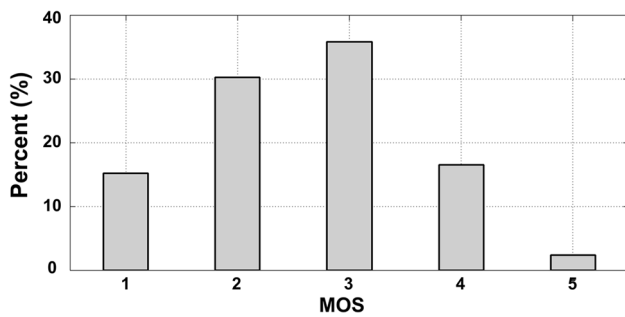
Fig. 6 Histogram of the obtained MOS (see in annex as recommended by the journal)

third choice respectively. Finally, 15.16 % of the observers said that the video is *bad* and 2.32 % *excellent*, respectively.

Observers consistently reacted strongly to high levels of loss and tended to avoid extreme scores (such as 5) for any video (the typical score is 3, *fair*), even if there was very little or no distortion compared to that of the original video [59]. However, the observers did not hesitate to rate a video extremely *poor* (score of 1) if the video was seriously impaired or the artifacts on the screen were barely visible. The information given here is important because it explains some of the results found in MultiQoE, where the subjects tend to evaluate the quality level of videos as *fair*.

Video QoE involves both application and user-oriented assessments. The viewer's individual interests, quality expectation, and service experience, are among the contributing factors to the perceived quality [25]. In general, human experience greater feelings of intensity in adverse situations than those that please them. In other words, observers are quick to criticize and slow to forgive. MOS takes less time to fall when distortions appear than to rise when distortions disappear [43]. Subjective quality depends significantly on where the lost packets are located and the nature of the related video content [28]. Since perception tests are time-consuming, costly, and unable to allow the quality to be assessed during real-time service operations, instrumental assessment methods are often preferred [27].

The test platform is a Desktop PC with Intel Core i5, 4 GB RAM, and a 21″ LCD monitor. The videos were played in the center of the monitor against a neutral gray background. A software was used to show the video sequences and collect the user scores. The distance from the eyes of the participants to the monitor was set at five times the height of the video sequence, as recommended by ITU. The seats were also adjusted so that the eyes of the participants were more likely to be at the same horizontal and vertical level as the center of the monitor.

The information available into the distorted video database (Component 3) will be used by Component 5

(MANN) to correlate the impact of network's impairments (informed by Component 4) on different video sequences from the HVS point-of-view.

3.4 Component 4: measurement model of factors affecting quality

MultiQoE uses the percentage of losses of the *I*, *P*, and *B* frames, the GoP structure, and video content features as input for the video quality estimator (Component 5—MANN). During the transmission of MPEG-4 videos, the dropping of packets (e.g., collisions, fading, or buffer overflow) carrying *I*, *P*, or *B* frames has a different visual impact on the delivered video. First of all, if the network discards one IP packet that contains the content of an *I* frame, the resulting distortions will be spread to all the dependent frames within the same GoP [60]. Second, if the content of the dropped packet belongs to a *P* frame, the impairments will spread to the remainder of the GoP. Third, if the dropped packet belongs to a *B* frame, the damage will only affect this frame. MultiQoE can be easily configured to estimate the quality level of videos encoded with different codecs and not only MPEG, where the system must be adapted to GoP structure and frame dependency of the codec.

Depending on the levels of spatial and temporal activities carried out in the video sequence, a GoP is composed of video frames with different sizes. For instance, video sequences with larger *I* frames (e.g., videos with high spatial complexity, such as *Mobile* video, like those displayed in Table 1) will be split up into several IP packets/frames to be transmitted over the network. Hence, the packet dropping probability of an *I* frame increases and has a different impact on the user's perception. The same process occurs in *P* and *B* frames for videos with high temporal complexity (e.g., the *Football* video). The GoP length has a strong influence in the composition of the MPEG flow as shown in Fig. 7. In the case of the GoP 10, the videos *Akiyo*, *Coastguard*, *Hall*, *Mother*, *News*, and *Silent* are mostly composed of *I* packets. The proportion of *P* and *B* frames dramatically increases when the GoP length is increased to 20 and 30. Owing to the difference in the visual impact of the *I*, *P*, and *B* frames, the increase of GoP length affects the degree of influence that the network impairments have on video perception.

Figure 8 demonstrates the PCC correlation between the selected parameters and MOS for the LST, MSHT, and HSMT clusters. In the case of the LST cluster, the highest correlation obtained is between MOS and the loss of an *I* frame (68.0 %). Thus, the results show that for videos with low levels of spatial and temporal complexity, the *I* frame has a higher impact on the user's perception. Figure 8 also reveals that the LST cluster has a high
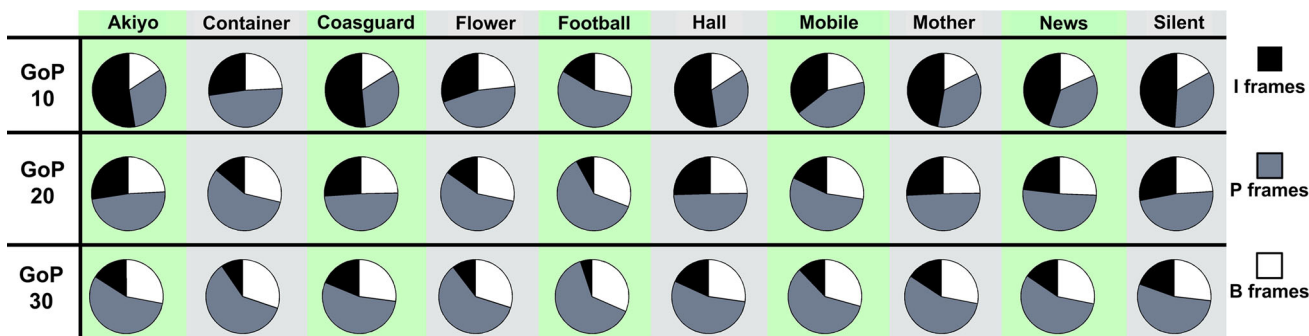
Fig. 7 The influence of the GoP length on the video stream composition (see in annex)
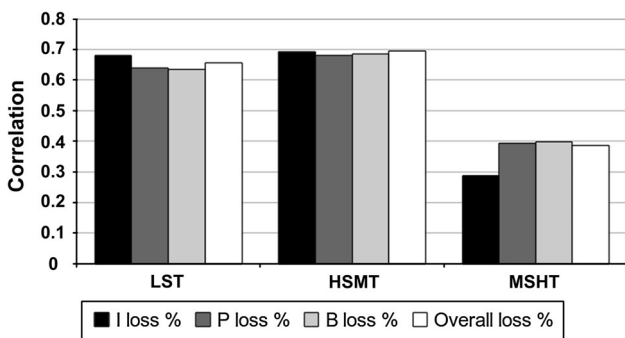

Fig. 8 The input correlation obtained with MOS (see in annex as recommended by the journal)

correlation between MOS and the selected parameters. Similarly in the case of the LST cluster, the HSMT cluster has a high level of correlation between MOS and the video/networks characteristics. On average, the MSHT cluster obtained a correlation of 68.82 %.

Figure 8 also demonstrates that the correlation values for the MSHT cluster are lower than HSMT and LST. This result can be explained by the fact that the MSHT cluster is composed of videos (*Coastguard* and *Football*) with the highest level of temporal activities. This leads to a higher number of *P* and *B* packets while the number of *I* packets remains the same. Hence, in terms of traffic composition, the MSHT cluster has a lower proportion of *I* packets for all the GoP lengths (see Fig. 7). This reduces the impact of the packets with an *I* frame that have to be dropped (as shown in Fig. 8). The results also show that observers tend to be more rigorous in their assessment of videos with a high level of spatial and temporal complexity.

3.5 Component 5: correlation of video characteristics, human experience, and network impairments into a predicted MOS

Component 5 is responsible for achieving a predicted/final MOS score by using a MANN to correlate video's characteristics, human's experience, and network's impairments into

MOS. MultiQoE performs well even with video flows not presented in the source video database. This is possible because MANN identifies patterns of video sequences (which can be different from the training flows) and provides an accurate prediction model in such scenarios. MultiQoE has been tested and validated as a dynamic and content-aware quality predictor to estimate the video quality of several types of video sequence in realistic WMNs, without any interaction with real viewers and with low complexity/processing.

MANNs have been used in many research areas to address problems that include function approximation, classification, and feature extraction, and allow complex tasks to be broken down into smaller and specialized tasks [61–63]. Each ANN is trained to become a specialist to carry out a specific task of the prediction system (e.g., for videos with a specific GoP length). Hence, it is possible to explore the advantages offered by MANNs in tackling problems that could not be solved with a single ANN. Moreover, the MANNs have a greater capacity for generalization, high performance, and providing an accurate prediction model.

There are many parameters that affect video quality level and their combined effect is unclear, but their relationships are thought to be non-linear [3]. ANNs can be used to learn this non-linear relationship that mimics the human perception of video quality. Thus, it is possible for ANNs to predict a pattern of sequences that they have been trained to deal with. The real challenge is to predict sequences that were not followed by the network in its training. With this goal in mind, the part of the videos that will be used for training should have the capacity to support the network with enough power to extrapolate patterns that may exist in other sequences [3], as demonstrated in MultiQoE results.

Our analytical studies found that the GoP length has a strong influence on the prediction of video quality, as illustrated in Fig. 7. In this context, the GoP length was selected as a key parameter and was divided into three specialized ANNs. Each ANN was programmed with a specific sub-database comprising GoP lengths of 10, 20,

and 30, to obtain better results. The reason for this is that each ANN is responsible for mapping the quality level of videos with a specific GoP length. Thus, each ANN has outputs designed for a particular GoP length and in the case of lengths between 10, 20, or 30, the final QoE estimate is calculated by one ANN.

An optimal linear combination of a weighted sum was constructed to yield a final result from the three outputs, because this method achieves a higher degree of accuracy than the single trained network. Another advantage is the adjustment for optimal weights during the tests and the training phases. The $MANN_{final}$ (generates a MOS as close as possible to human scores) output is performed by:

$$MANN_{final} = \sum_{j=1}^{p} \alpha_j y_j$$

where j is the number of ANNs. α is 0 or 1. y: is the ANN output. $\alpha_j y_j$: is the output-product of the j-th ANN. p: is the number of ANNs composing the MANN. When the GoP length of the incoming flow matches with a particular ANN GoP length, its weight is 1 and the weights of the other ANN GoPs are 0.

3.6 Implementation: use case in MPEG and SDN/OpenFlow systems

MultiQoE uses information about the video's characteristics, network's impairments, human's experience, and a MANN to predict the video MOS. MultiQoE collects the video characteristics and network impairments by using a deep packet inspector module. Video coding standards (e.g., MPEG) specify the bitstream format and the decoding process in a video sequence. Each flow starts with a sequence header, followed by a GoP header, and then by one or more coded frames. Each IP packet contains one or more video frames.

The deep packet inspector examines the MPEG bitstreaming and verifies which frame was lost in a GoP, without decoding the video payload. The packet inspector also collects information about the frame type and intra-frame dependency, which are described in the video sequence and GoP headers. MultiQoE uses a correlation between DCT coefficients, MV, and frame size to define the level of spatial and temporal video's characteristics. To reduce the computational cost of deep packet inspection schemes, it is expected that in a near future the video codec will provide additional information on the encoded flows by using a video descriptor scheme (e.g., as proposed in the recent MPEG DASH standard, named Media Presentation Description (MPD)) to allow cross-layer multimedia networking solutions to improve the usage of network resources and the user perception.

For the MANN training phase, MultiQoE collects the experience of real subjects (e.g., MOS) by using an off-line approach, where it asks a panel of viewers to grade the quality level of distorted videos. Since the MultiQoE MANN is trained with a set of video's sequences, network's impairments, and MOS before its use, it will only use information about the frame type and the percentage of losses of *I*, *P*, and *B* frames of a GoP during its in-service video estimator procedures, which aims to reduce the processing time. MultiQoE can introduce an extra delay to predict the video quality level with high accuracy, but advanced filter, classification, and optimization techniques, as well as cloud computing techniques can be used to improve the system performance.

A simple example of the implementation of MultiQoE in real scenarios is described as follows. MultiQoE agents can be configured in network devices (e.g., together with routers or access points) to estimate the quality of video streams even when they have different encoding patterns, genres, content types, and packet loss rates. In our future work, with the help of Software Defined Networks (SDN), such as OpenFlow, MultiQoE will be installed in SDN/OpenFlow environments and orchestrate all of its control modules/components in production systems. For instance, MultiQoE can be placed as an external application module and interacts with the OpenFlow controller in a network, by using JavaScript Object Notation (JSON) or another OpenFlow *Application Programming Interface* (API).

The OpenFlow protocol can capture information about all video flows from the OpenFlow switches/routers in a network and pass them to the MultiQoE application via an OpenFlow controller, including information about packet loss and frame type as proposed in a previous work [64].

Following the OpenFlow architecture, when a *Packet_in* from a video packet is received by the controller, it sends *flowmod* commands to the switches instructing them to send the video packets to its destination and, at the same time, a copy of the packets are forwarded to the MultiQoE application. The packet inspector scheme connected to OpenFlow switches/routers (or even linked to the controller as proposed in [65]) examines the packet header/MPEG bit streaming and gets information about frame type and size. Upon receiving all networking statistics from the OpenFlow controller and (video) packet/header information from the packet inspector, MultiQoE is able to trigger its QoE video quality prediction system and provides a predicted MOS for each flow.

MultiQoE does not need the original video sequence to estimate the video quality level, which reduces the computational complexity and, at the same time, opens up the possibilities of the video quality prediction deployment. To improve the MultiQoE performance, the system is trained with a large set of video sequences and network

impairments before its use as recommended in [66]. When MultiQoE is triggered, it starts its prediction procedures without any interaction with real viewers, but with high accuracy as presented in Sect. 4.

4 Performance evaluation

This section demonstrates the benefits of MultiQoE in a practical environment. This was achieved by employing a use case in an IEEE 802.11-based WMN system, which was implemented together with MultiQoE, to measure the quality level of real video sequences distributed in WMNs. The efficiency of MultiQoE is compared to widely-used QoE metrics such as PSNR, VQM and SSIM, as well as, PSQA and MOS collected from real observers. Our results also rely on two key estimation methods, namely Pearson Correlation Coefficient (PCC) [67] and Mean Squared Error (MSE) [31] as recommended by the VQEG [68, 69].

The simulated scenario uses the topology and characteristics of the WMN backbone installed at the UFPA campus in the Amazon/Brazil. The setting consists of several buildings interspersed with parking areas and woodland. The topology is composed of 6 mesh routers, 2 of which are gateways, as depicted in Fig. 4. In addition, a mesh client was positioned to receive the video streaming from either Gateway 1 (G1) or Gateway 2 (G2). The client experienced different packet loss rates because during the tests the user location and wireless resource conditions/impairments were changed at random (up to 90 % of network congestion—see Fig. 5). Ten widely used Internet-based videos were chosen with different patterns and characteristics (duration, content types, and GoP length) as explained in Sect. 3.1.

The experiments were carried out by using Network Simulator 2.34 [70], Evalvid Tool [71], MSU Video Quality Measurement Tool [72], and the MANN was built with the aid of Matlab. The distorted video database was split into two subsets: one for training and another one for testing the generalization performance of the trained system. With this procedure, it is possible to make sure that the set of videos of the training set and test set come from disjoint video sequences with quite different video content. Each selected video was simulated 90 times (by varying the network packet loss rate and GoP lengths—10, 20, and 30) to provide a large video database and this resulted in a total of 900 videos.[1] 810 and 90 videos were randomly selected from this database for the training and test databases, respectively. Table 3 outlines other parameters used in the experiments.

---

[1] All videos used in the experiments are available in https://www.youtube.com/channel/UCTp9d3EYjoERYbM5pMfImog.

Table 3 Simulation parameters configured in the simulator

| Parameter	Value
Frequency	2.4 GHz
IEEE model	IEEE 802.11 g
Anttena type	Omnidirectional
Anttena gain	12 dB
Anttena height	8 m
Video frame rate	25 frames/s
Packet loss model	Gilbert-Elliot
Shadowing deviation	4.4
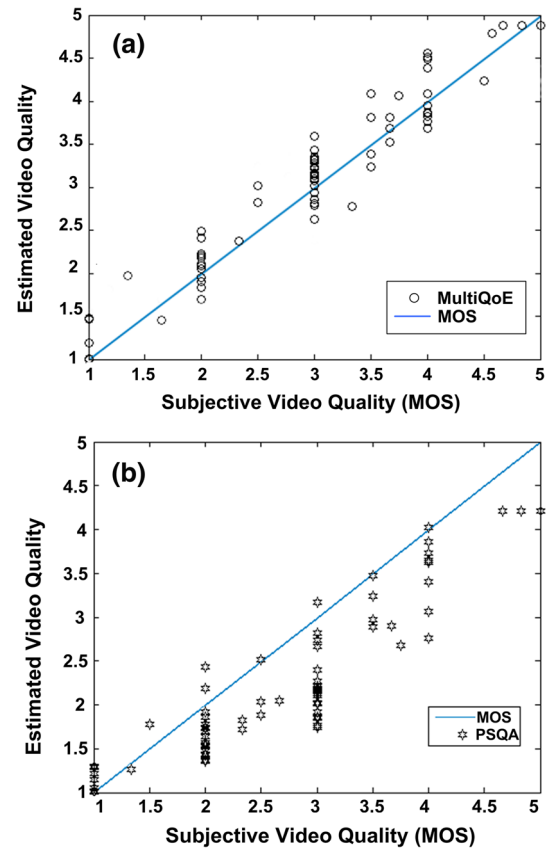
Fig. 9 MultiQoE versus PSQA versus MOS. **a** MultiQoE versus MOS, **b** PSQA versus MOS (see in annex)

While conducting the subjective evaluation tests, we followed the ITU-T MOS recommendations (with 55 observers) to obtain accurate results. The observers included undergraduates, post-graduate students, and university staff. The test platform is the same as that described in Sect. 3.4.
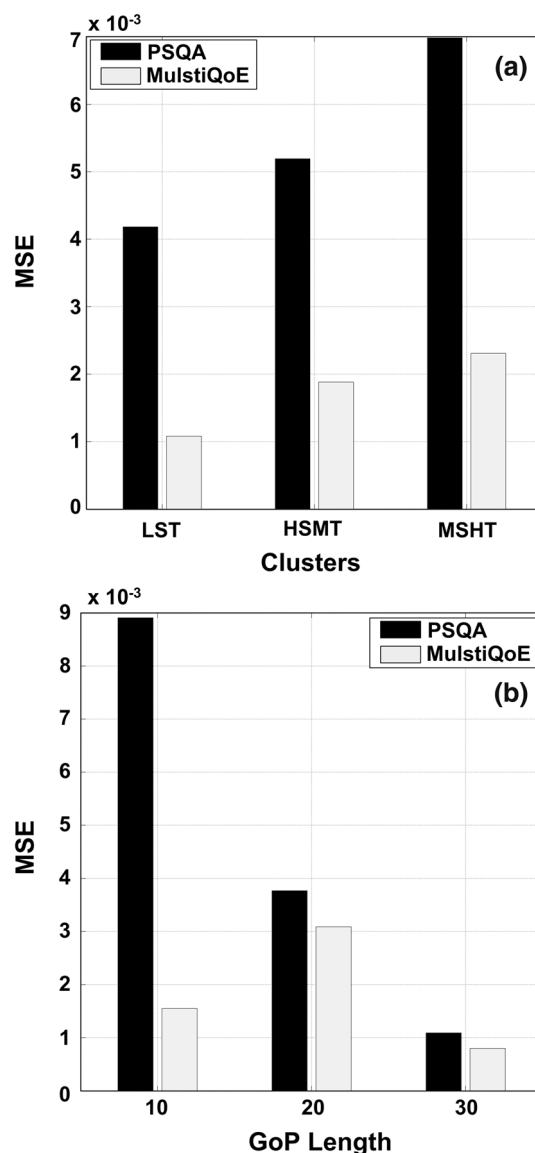
Figure 9a shows the results obtained when the WMN is configured with MultiQoE. It can be observed that MultiQoE and MOS achieve similar scores, while PSQA does not correlate well with MOS (Fig. 9b). The PSQA values produce scores not close to the MOS line because the parameters used as input (losses in I, P, and B frames and
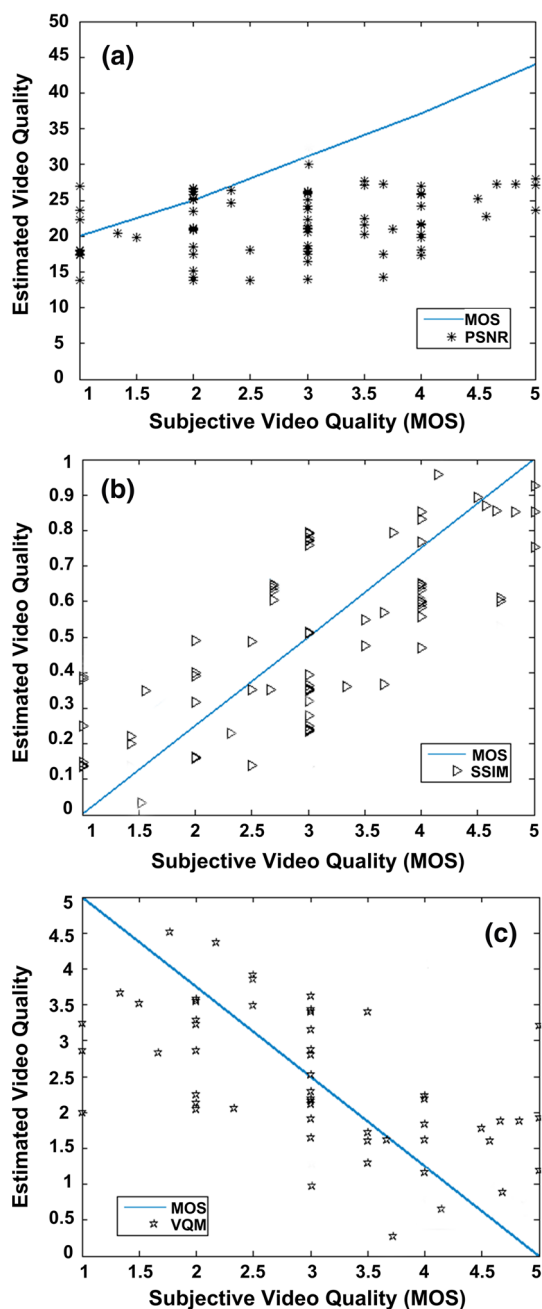
Fig. 10 Video quality level vs. MOS. **a** PSNR versus MOS, **b** SSIM versus MOS, **c** VQM versus MOS

Fig. 11 MSE for different clusters and GoP length. **a** MSE for each cluster, **b** MSE for each GoP length

GoP length) and the correlation model based on a RNN, are not enough to predict the quality level of the videos accurately. Moreover, in contrast to PSQA, MultiQoE improves the procedures for estimating video quality by using the GoP length combined with a specific ANN (and the spatial/temporal (clustered) activities of the videos) as input for its correlation component. The results reveal that videos with varying content features have a different impact on the user's perception even when the wireless channel conditions are similar.

Figure 10a–c show the results of PSNR, SSIM, and VQM, respectively, when compared with MOS. It is evident that there is a poor correlation between the objective metrics and MOS scores and that this does not reflect the user's perception.

Further results show the benefits of MultiQoE in predicting the quality level of videos, by measuring the MSE values of both MultiQoE and PSQA for each type of cluster and GoP length (as shown in Fig. 11a, b). This analysis is important since it reveals the performance of QoE estimation methods for video clusters in scenarios with different congestion levels and packet loss rates. It is worth noting that Fig. 11a shows that MultiQoE has the lowest error, with only 1.08×10^{-3}, for the LST cluster, while PSQA has 4.18×10^{-3}. It should also be pointed out that
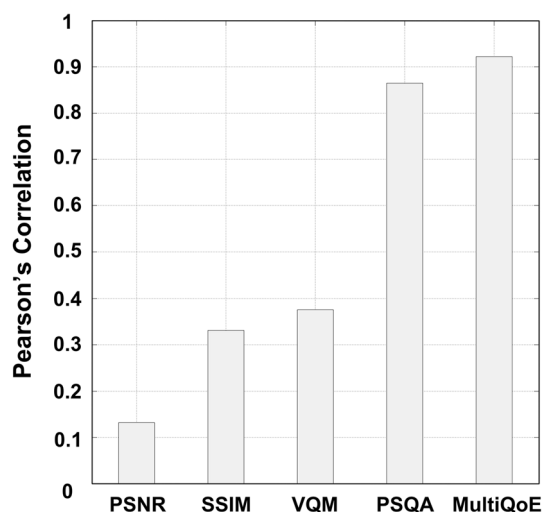
Fig. 12 Pearson correlation for whole proposal (see in annex as recommended by the journal)

MultiQoE has the best performance for clusters HSMT and MSHT with 1.88×10^{-3} and 2.31×10^{-3} compared with PSQA (5.19×10^{-3} and 6.98×10^{-3}), respectively. Distortions in foreground areas, such as human faces in high motion and complexity videos, caused lower subjective ratings, while similar artifacts in the background went unnoticed for videos with low motion and complexity levels.

Figure 11b illustrates the performance of the MultiQoE and PSQA for all GoP lengths. In the case of GoP 10, the MSE results for MultiQoE and PSQA are 1.55×10^{-3} and 8.91×10^{-3}, respectively. For GoP 20, MultiQoE presents a MSE of 3.09×10^{-3}, while PSQA only 3.77×10^{-3}. Finally, for videos encoded with GoP 30, the MultiQoE MSE is only 0.081×10^{-3}, while PSQA is 1.09×10^{-3}. For GoP 30, PSQA shows a better performance than PSQA, because the drop of I frames had a great impact on the video quality level. According to the MSE results, on average, MultiQoE improves the accuracy of the system by 32.22 % when compared with PSQA.

Figure 12 illustrates the PCC values for MultiQoE, PSQA, PSNR, SSIM, and VQM, where 1 indicates a perfect match between the predicted measurements and the subjective ratings and 0 indicates no correlation. The PCC coefficient obtained by MultiQoE is 0.922 (an improvement of 7 % compared with PSQA). When the system is configured to analyze the quality level of videos based only on PSNR, SSIM, and VQM, the PCC values are 0.132, 0.331, and 0.376, respectively (Fig. 12). The results confirm that the objective metrics achieved are poor compared with MOS, PSQA, and MultiQoE.

It can be observed that MultiQoE causes a low number of errors with subjective ratings for a variety of videos ranging from low activity such as *Akiyo* and *News* to high

activity sequences, such as *Football* and *Coastguard*. MultiQoE also produces good results for video sequences that combine both low-activity and high-activity scenes, such as *Silent* and *Flower*. This is because MultiQoE uses the GoP length and the spatial/temporal (clustered) activities of the videos as input for its correlation component.

After exploring the impact of all MultiQoE components in multimedia wireless systems, we highlight the accuracy of the video quality estimator in assessing the quality level of real-time video sequences not (previously) included in the video database. New experiments were carried out in the same simulation scenarios (congestion levels and number of repetitions—95 % confidence interval), where one new MPEG4 video called *Grandma* was used in the simulations.

On average, the PCC result obtained by MultiQoE is 0.892 for the *Grandma* flow. When the system is configured to analyze the quality level of videos based only on PSNR, SSIM, VQM, and PSQA, the PCC values are 0.124, 0.348, 0.416 and 0.821, respectively. The results confirm that objective metrics perform poorly compared with those of MOS, PSQA, and MultiQoE. When *Grandma* is included in the source database and trained off-line, the PCC result is of 0.928. This is possible because, in addition to the benefits of MANN in identifying patterns of video sequences, which they were trained to deal with (as happened with *Mother* and *Flower*), and providing an accurate prediction model, MultiQoE uses a set of feed-forward back-propagation networks that are supplied with subjective MOS scores. These parameters enable MultiQoE to measure the quality level of videos even when they have different encoding patterns, content types, and network impairments/errors/congestions.

5 Conclusion

The evolution of wireless access technologies, services, and protocols has created a plethora of new human-centric environments featuring an ever-increasing amount of wireless devices and multimedia content. QoE assessment and control solutions allow network providers to keep and attract new customers, while optimizing network resource and enlarging their portfolios. Therefore, parametric in-service QoE assessment models are needed to ensure the success of multimedia wireless networks and have attracted a lot of attention from both academia and industry. MultiQoE provides a modular and non-intrusive video quality estimator implemented over wireless mesh systems that can be easily adapted to networks with different underlying technologies. MultiQoE works without the need for any decoding which saves time and reduces processing.

Our experiments investigated the relationship between video's characteristics, human's experience, network's impairments, and predicted MOS scores. On average, the results show that MultiQoE has a high correlation with MOS (0.922) and a low MSE 1.75×10^{-3}, while PSQA produced 5.45×10^{-3}. The benefits of MultiQoE have also been highlighted when compared with PSNR, VQM, and SSIM. MultiQoE can be used together with optimization and management schemes to improve the usage of network resources, as well as system performance in key networking areas, such as pricing, routing, or mobility.

In future works, large-scale experiments will be conducted to investigate the impact of the proposed solution in networks with many users and a large set of video flows. Thus, it will be possible to analyze the MultiQoE computation cost to predict the quality level of user and content-generated video sequences over wireless networks with different channel modules and errors. An OpenFlow prototype with Cloud Computing support is being developed to evaluate the benefit of MultiQoE in production multimedia networks.

References

1. Cerqueira, E., Zeadally, S., Leszczuk, M., Curado, M., & Mauthe, A. (2011). Recent advances in multimedia networking. *Multimedia Tools and Applications, 54*(3), 635–647.
2. Mehmood, R., Alturki, R., & Zeadally, S. (2011). Multimedia applications over metropolitan area networks (MANs). *Journal of Network and Computer Applications, 34*(5), 1518–1529.
3. Khan, A., Sun, L., & Ifeachor, E. (2010). Learning models for video quality prediction over wireless local area network and universal mobile telecommunication system networks. *IET Communications, 4*(12), 1389–1403.
4. Chaari, H., Mnif, K., & Kamoun, L. (2012). An overview of quality assessment methods of video transmission over wireless networks. In *Proceedings of 16th IEEE mediterranean electrotechnical conference,* (pp. 741–744).
5. Greengrass, J., Evans, J., & Begen, A. C. (2009). Not all packets are equal, part I: Streaming video coding and sla requirements. *IEEE Internet Computing, 13*(1), 70–75.
6. Benyamina, D., Hafid, A., & Gendreau, M. (2012). Wireless mesh networks design: A survey. *IEEE Communications Surveys & Tutorials, 14*(2), 299–310.
7. Sarisaray-Boluk, P. (2013). Performance comparisons of the image quality evaluation techniques in wireless multimedia sensor networks. *Wireless Networks, 19*(4), 443–460.
8. Qiu, X., Liu, H., Ghosal, D., Mukherjee, B., Benko, J., Li, W., et al. (2011). Enhancing the performance of video streaming in wireless mesh networks. *Wireless Personal Communications, 56*(3), 535–557.
9. Chikkerur, S., Sundaram, V., Reisslein, M., & Karam, L. J. (2011). Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting, 57*(2), 165–182.
10. Winkler, S. (2009). Video quality measurement standards: Current status and trends. In *7th International conference on information, communications and signal processing, (ICICS), December 2009* (pp. 1–5). Macau.
11. Aguiar, E. S., Riker, A., Abelem, A., Cerqueira, E, & Mu, M. (2012). Video quality estimator for wireless mesh networks. *IEEE/ACM 20th International workshop on quality of service* (IEEE/ACM IWQoS 2012), 1–9.
12. Park, Y., Seok, Y., Choi, N., Choi, Y., & Bonnin, J. M. (2006). Rate-adaptive multimedia multicasting over IEEE 802.11 wireless LANs. In *3rd IEEE consumer communications and networking conference, (CCNC).* doi:10.1109/CCNC.2206.1593011.
13. Revés, X., Nafisi, N., Ferrús, R., & Gelonch, A. (2006). User perceived quality evaluation in a b3g network testbed. In *Proceedings of IST mobile summit. Mykonos, Greek.*
14. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.
15. Mohamed, S., & Rubino, G. (2002). A study of real-time packet video quality using random neural networks. *IEEE Transactions on Circuits and Systems for Video Technology, 12*(12), 1071–1083.
16. Recommendation, I. T. U. R. B. T.500-11. (2002). Methodology for the subjective assessment of the quality of television pictures. *International telecommunication union.* Geneva, Switzerland.
17. Recommendation, I. T. U. T. P. 800. (1996). Methods for subjective determination of transmission quality. *International telecommunication union,* Geneva, Switzerland.
18. Moorthy, A. K., Seshadrinathan, K., Soundararajan, R., & Bovik, A. C. (2010). Wireless video quality assessment: A study of subjective scores and objective algorithms. *IEEE Transactions on Circuits and Systems for Video Technology, 20*(4), 587–599.
19. Farias, M. C., & Carvalho, M. M. (2010). Video quality assessment based on data hiding for IEEE 802.11 wireless networks. In *IEEE International symposium on broadband multimedia systems and broadcasting (BMSB)* (pp. 1–6). Shanghai, China.
20. Recommendation, I. T. U. T. P. 910. (1999). Subjective video quality assessment methods for multimedia applications. *International telecommunication union.* Geneva, Switzerland.
21. Chen, K. T., Tu, C. C., & Xiao, W. C. (2009). OneClick: A framework for measuring network quality of experience. In *IEEE INFOCOM 2009* (pp. 702–710). Rio de Janeiro, Brazil.
22. Raake, A., & Möller, S. (2011). Recent multimedia qoe standardization activities in itu-t sg12. *IEEE COMSOC MMTC E-letter-special issue on quality of experience issues in media delivery* (pp. 10–14).
23. I. T. U T. SG12. (2010). Draft terms of reference (ToR) for P.NAMS. Geneva, Switzerland. ITU. http://www.itu.int/md/meetingdoc.asp?lang=en&parent=T09-SG12-091103-TD-GEN-0146.
24. I. T. U T. SG12. (2009). Draft terms of reference (ToR) for P.NBAMS. Geneva, Switzerland. ITU. http://www.itu.int/md/T09-SG12-110118-TD-GEN-0521.
25. Liao, N., & Chen, Z. (2010). A packet-layer video quality assessment model based on spatiotemporal complexity estimation. In *International society for optics and photonics on visual communications and image processing,* August 2010. doi:10.1117/12.863473.
26. Yamagishi, K., & Hayashi, T. (2008). Parametric packet-layer model for monitoring video quality of IPTV services. In *IEEE International conference on communications (ICC), May 2008.* (pp. 110–114). Beijing, China.
27. Garcia, M. N., & Raake, A. (2011). Frame-layer packet-based parametric video quality model for encrypted video in IPTV services. In *Third international workshop on quality of multimedia experience (QoMEX)* (pp. 102–106). Mechelen, Belgium.

28. Yang, F., Wan, S., Xie, Q., & Wu, H. R. (2010). No-reference quality assessment for networked video via primary analysis of bit stream. *IEEE Transactions on Circuits and Systems for Video Technology, 20*(11), 1544–1554.

29. Javadi, F., & Jamalipour, A. (2011). A multi-path cognitive resource management mechanism for QoS provisioning in wireless mesh networks. *Wireless Networks, 17*(1), 277–290.

30. Urrea Duque, J. P., & Gaviria Gomez, N. (2012). Quality assessment for video streaming P2P application over wireless mesh network. In *XVII Symposium of image, signal processing, and artificial vision (STSIVA), September 2012.* (pp. 99–103). Colombia.

31. Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine, 26*(1), 98–117.

32. Xiao, F. (2000). DCT-based video quality evaluation. *Final Project for EE392 J*, 769.

33. Ghareeb, M., & Viho, C. (2010). Hybrid qoe assessment is well-suited for multiple description coding video streaming in overlay networks. In *Eighth annual communication networks and services research conference (CNSR), May 2010.* (pp. 327–333). Montreal, Canada.

34. Sagatov, E. S., Sukhov, A. M., & Calyam, P. (2010). Influence of distortions of key frames on video transfer in wireless networks. In *5th International symposium on I/V communications and mobile network (ISVC), September 2010.* (pp. 1–4). Rabat, Morocco.

35. Khan, A., Sun, L., & Ifeachor, E. (2009). Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks. In *IEEE International conference on communications (ICC), June 2009.* (pp. 1–5). Dresden, Germany.

36. Khan, A., Sun, L., Jammeh, E., & Ifeachor, E. (2010). Quality of experience-driven adaptation scheme for video applications over wireless networks. *IET Communications, 4*(11), 1337–1347.

37. Khan, A., Sun, L., & Ifeachor, E. (2009). Impact of video content on video quality for video over wireless networks. In *Fifth international conference on autonomic and autonomous systems (ICAS). April 2009.* (pp. 277–282). Valencia, Spain.

38. Orellana-Romero, E., SanMartin-Hernandez, J., Duran-Faundez, C., Lecuire, V., & Aguilera, C. (2011). Sim-LIT: A simulation framework for image quality assessment in wireless visual sensor networks under packet loss conditions. In *XXX International conference of the chilean computer science society (SCCC).* Curico, Chile.

39. Assegie, S., ZareAfifi, S., Verma, R., King, B., Salama, P., & Kim, D. (2012). Quality assessment tests to be used in a wireless image sensor network. In *IEEE International conference on electro/information technology (EIT), May 2012.* (pp. 1–6). Indianapolis, USA.

40. Khashei, M., Zeinal Hamadani, A., & Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications, 39*(3), 2606–2620.

41. Bouzerdoum, A., Havstad, A., & Beghdadi, A. (2004). Image quality assessment using a neural network approach. In *Proceedings of the fourth IEEE international symposium on signal processing and information technology, December 2004.* (pp. 330–333).

42. Gastaldo, P., Rovetta, S., & Zunino, R. (2002). Objective quality assessment of MPEG-2 video streams by using CBP neural networks. *IEEE Transactions on Neural Networks, 13*(4), 939–947.

43. Le Callet, P., Viard-Gaudin, C., & Barba, D. (2006). A convolutional neural network approach for objective video quality assessment. *IEEE Transactions on Neural Networks, 17*(5), 1316–1327.

44. Carrai, P., Heynderickz, I., Gastaldo, P., & Zunino, R. (2002). Image quality assessment by using neural networks. In *IEEE International symposium on circuits and systems (ISCAS).* (Vol. 5, pp. 253–256). Scottsdale, USA.

45. Narwaria, M., & Lin, W. (2012). SVD-based quality metric for image and video using machine learning. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics, 42*(2), 347–364.

46. Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks: A review. *Pattern Recognition, 35*(10), 2279–2301.

47. Mohamed, S., Rubino, G., Cervantes, F., & Afifi, H. (2001). Real-time video quality assessment in packet networks: A neural network model. *Institut National de Recherche en Informatique et Automatique.* http://hal.archives-ouvertes.fr/inria-00072437/.

48. Frank, P., & Incera, J. (2006). A neural network based test bed for evaluating the quality of video streams in IP networks. In *Electronics, robotics and automotive mechanics conference, September 2006.* (1, pp. 178–183). Cuernavaca, Mexico.

49. Greengrass, J., Evans, J., & Begen, A. C. (2009). Not all packets are equal, part 2: The impact of network packet loss on video quality. *IEEE Internet Computing, 13*(2), 74–82.

50. Moorthy, A. K., & Bovik, A. C. (2011). Visual quality assessment algorithms: What does the future hold? *Multimedia Tools and Applications, 51*(2), 675–696.

51. Video Trace Library. National Science Foundation. http://trace.eas.asu.edu/.

52. Seeling, P., Fitzek, F. H., & Reisslein, M. (2007). *Video traces for network performance evaluation: A comprehensive overview and guide on video traces and their utilization in networking research.* Dordrecht: Springer.

53. Mitchell, J. L., Pennebaker, W. B., Fogg, C., & LeGall, D. J. (1996). *MPEG video compression standard.* Norwell: Kluwer Academic Publishers.

54. Ries, M., Crespi, C., Nemethova, O., & Rupp, M. (2007). Content based video quality estimation for H. 264/AVC video streaming. In *Wireless communications and networking conference (WCNC)* (pp. 2668–2673). Hong Kong.

55. Passos, D., Teixeira, D. V., Muchaluat-Saade, D. C., Magalhães, L. C. S., & Albuquerque, C. (2006). Mesh network performance measurements. In *International information and telecommunications technologies symposium (I2TS)* (pp. 48–55). Cuiaba, Brazil.

56. Yamada, T., Yachida, S., Senda, Y., & Serizawa, M. (2010). Accurate video-quality estimation without video decoding. In *IEEE International conference on acoustics speech and signal processing (ICASSP)* (pp. 2426–2429). Dallas, USA.

57. Recommendation, I. T. U. R. P. 911. (1998). Subjective audio-visual quality assessment methods for multimedia applications. *International telecommunication union.* Geneva, Switzerland.

58. Seshadrinathan, K., Soundararajan, R., Bovik, A. C., & Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing, 19*(6), 1427–1441.

59. Venkataraman, M., & Chatterjee, M. (2011). Inferring video QoE in real time. *IEEE Network, 25*(1), 4–13.

60. Wan, S., Yang, F., & Xie, Z. (2010). Evaluation of video quality degradation due to packet loss. In *IEEE International symposium on intelligent signal processing and communication systems (ISPACS), December 2010.* (pp. 1–4). Chengdu, China.

61. Zheng, Y., & Meng, Y. (2011). Modular neural networks for multi-class object recognition. In *IEEE International conference on robotics and automation (ICRA)* (pp. 2927–2932). Shanghai, China.

62. Choi, H., & Lee, C. (2011). Motion adaptive deinterlacing with modular neural networks. *IEEE Transactions on Circuits and Systems for Video Technology, 21*(6), 844–849.

63. Gradojevic, N., Gençay, R., & Kukolj, D. (2009). Option pricing with modular neural networks. *IEEE Transactions on Neural Networks, 20*(4), 626–637.
64. Georgopoulos, P., Elkhatib, Y., Broadbent, M., Mu, M., & Race, N. (2013). Towards network-wide QoE fairness using openflow-assisted adaptive video streaming. The ACM SIGCOMM 2013 workshop on future human-centric multimedia networking. Hong Kong.
65. Qosmos. (2013). ixEngine: Embedding DPI and L7 intelligence in an SDN architecture. Site.
66. Goudarzi, P. (2013). A no-reference low-complexity QoE measurement algorithm for H264 video transmission systems. *Scientia Iranica D, 20*(3), 721–729.
67. Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). *Noise reduction in speech processing*. Heidelberg: Springer.
68. Video Quality Experts Group. (2000). Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase I. *VQEG*.
69. Video Quality Experts Group. (2003). Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II. *VQEG*.
70. Issariyakul, T., & Hossain, E. (2011). *Introduction to network simulator NS2*. New York: Springer.
71. Klaue, J., Rathke, B., & Wolisz, A. (2003). Evalvid: A framework for video transmission and quality evaluation. In P. Kemper & W. H. Sanders (Eds.), *Computer performance evaluation. Modelling techniques and tools*. Berlin: Springer.
72. Vatolin, D., Moskvin, A., Petrov, O., & Trunichkin, N. (2009). *Msu video quality measurement tool*. http://www.download3k.com/search/MSU-Video-Quality-Measurement-Tool.

Elisangela Aguiar is a Ph.D. student at Federal University of Pará (UFPA) in Brazil. Her interests include wireless and mobile networks, multimedia, routing, quality of service, and quality of experience.

André Riker is Ph.D student at University of Coimbra / Department of Informatics Engineering. His research involves energy efficiency of Machine-to-Machine communication, video transmission and Quality of Experience.

Eduardo Cerqueira received his Ph.D. in Informatics Engineering from the University of Coimbra, Portugal, in 2008. He is an associate professor at the Faculty of Computer Engineering and Telecommunications of the UFPA in Brazil. His publications include four books, four patents and over than 100 papers in national/international refereed journals/conferences. He is involved in the organization of several international conferences and workshops, including Future Multimedia Networking, Latin America Conference on Communications, and Latin American Conference on Networking. He has been serving as a Guest Editor for many special issues of various peer-reviewed scholarly journals. His research involves multimedia, Future network, Quality of Experience, Mobility.

Antônio Abelém is Associate Professor on the Computer Science Faculty at the Federal University of Pará (UFPA) in Belém, capital of the Brazilian state of Pará. He holds a B.Sc. degree in Electrical Engineering from the same University, an M.Sc. degree in Computer Systems from the Catholic University of Rio de Janeiro (PUC-Rio) and a Ph.D. in Informatics also from PUC-Rio. His research interests include future Internet, optical networks, ad-hoc networks, multicast, QoS and QoE.

Mu Mu received his Master's degree of Science from Darmstadt University of Technology, Germany in 2006. In 2007, He joined the School of Computing and Communications at Lancaster University, where he leads the research on the topic of quality of experience (QoE). His degree of Ph.D. in Computer Science was awarded by Lancaster University. As a Research Associate at Lancaster University, Lancaster, Dr. Mu has been a named researcher of several European and UK projects. His publications appeared in top-rank conferences and journals. His current research interests include user experience in multimedia services, data mining social media, and user-centric content networking.

Torsten Braun got his Ph.D. degree from University of Karlsruhe (Germany) in 1993. From 1994 to 1995 he has been a guest scientist at INRIA Sophia-Antipolis (France). From 1995 to 1997 he has been working at the IBM European Networking Centre Heidelberg (Germany) as a project leader and senior consultant. He has been a full professor of Computer Science at the University of Bern (Switzerland) since 1998 and director of the Institute of Computer Science and Applied Mathematics at University of Bern from 2007 to 2011. He has been member of the SWITCH (Swiss education and research network) board of trustees since 2001. He has over than 150 papers in national/international refereed journals/conferences.

Marilia Curado is an Assistant Professor at the Department of Informatics Engineering of the University of Coimbra, Portugal, from where she got a Ph.D. in 2005. Her research interests are multimedia, QoE, Mobility, Routing, and Resilience. She has over than 100 papers in national/international refereed journals/conferences. She has participated in several national projects, in Networks of Excellence from IST FP5 and FP6, in the IST FP6 Integrated Projects, EuQoS and WEIRD, and on ICT FP7 STREPs MICIE and GINSENG.

Sherali Zeadally received his Bachelor degree in Computer Science from University of Cambridge, England, and the Doctoral degree in Computer Science from University of Buckingham, England, in 1996. He is an Associate Professor at University of Kentucky, USA. He currently serves on the Editorial Boards of 15 peer-reviewed international journals. He has been serving as a Guest Editor for over a dozen special issues of various peer-reviewed scholarly journals. He is a Fellow of the British Computer Society and a Fellow of the Institution of Engineering Technology, UK. His research interests include computer networks including wired and wireless networks, network and system security, mobile computing, ubiquitous computing, RFID, multimedia, performance evaluation of systems, and networks.