Chapter 3

Multimedia Human-Centric Networking: Concepts, Technologies and Trends

Eduardo Cerqueira, Aldri Santos, Denis Rosário, Torsten Braun, and Mario Gerla

Abstract

The transmission of multimedia content will represent up to 90% of all Internet traffic in a few years, where it will be mainly accessed over wireless networks. However, this new wireless multimedia era requires instantaneous adaptation of multimedia content and the network resources according to user's preferences, experiences, or interests. In this context, human-centric multimedia networking (HCMN) appears as a promising model for next generation of wireless multimedia networks. In such scenarios, users will produce, share, and consume video flows ubiquitously, ranging from entertaining to real-time video flows about natural disasters or surveillance. Although, optimizations in HCMN scenarios must consider issues related to the network, the video characteristics, and, especially, human's preferences or the human's visual system. In this way, HCMN systems place the human's experience in the centre of mobile video services, protocols, and applications, where the video transmission process must be done and optimized in real-time according to the human's perceptions, content's characteristics, and also context-awareness. In this work, we introduce the basic concepts for the video transmission in HCMN systems. We also detail main existing mechanisms, which aim to improve the performance of HCMN system in sharing video content, including Quality of Experience (QoE)-based solutions for handover, routing, error correction, decision-making, and controlling the dissemination of video flows in wireless multimedia-aware environments.

3.1. Introduction

The evolution of wireless access technologies, mobile devices, and protocols together with the constant demand for video applications has created new human-centric multimedia networking (HCMN) environments [Lu et al. 2011, Cerqueira et al. 2014]. This is due to the use of multimedia services over wireless networks is becoming part of our personal and professional lives, which allows new video sharing experiences for millions of people in the worldwide. In this context, humans are changing their traditional communication paradigms based on voice calls or text messages to real-time video calls or sharing rich human digital memories in social networks. Many of our life moments can be delivered digitally. For instance, we can feel that we are in two places at once, by using face-to-face video calling services, such as provided by Apple FaceTime, Tango, Vonage, and others. Moreover, the real-time multimedia transmissions to thousands of people over wireless networks will take place in the coming big events, such as in the FIFA World Cup and the Olympic Games in Brazil in 2014 and 2016, respectively.

In addition to entertainment, calling, and human digital memory services, life videos are also important for disaster and surveillance services [Miao et al. 2012]. Live streaming video flows provide users and authorities (e.g., firefighters and paramedics) with more precise information than simple text messages, and also allow them to determine a suitable action based on rich visual information, while reducing human reaction times. For instance, vehicles or Unmanned Aerial Vehicles (UAVs) can cooperate with each other to disseminate short videos of dangerous situations, which are hundred or thousand of meters ahead to visually inform drivers and rescue teams about them.

Motivated by the great success and importance of video-based applications and services for the IT society, the multimedia market will continue to expand during the next years with real-time video services for mobile devices. For instance, CISCO projected that over 10 billion devices will be in used until 2016, and 71% of all mobile data traffic is expected to be videos by that time as well. Moreover, the multimedia transmission will represent up to 90% of the global IP data traffic in a few years [Cisco 2012]. Hence, the multimedia industry believes that mobile video consumption is growing at such a fast pace, that it is comparable to an unstoppable train, and thus get on board or get knocked down!

The video coded/decoded (CODEC) [Hanzo et al. 2007] plays an important role in HCMN environments, where a set of video CODEC have been used to transmit videos in the Internet, including Google VP9 [Sharabayko et al. 2013], H.264/MPEG-4 AVC [Puri et al. 2004], MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [Sodagar 2011], and others. In general, real-time multimedia traffic consists of one or more media streams with different spatial or temporal (motion and complexity level) video activities and features. From the aspect of video characteristics, a typical hierarchical-based compressed video, such as H.264 or MPEG-4, is composed of three types of frames, namely I- (Intra-coded), P- (Predictive-coded), and B- (Bidirectionally predictive-coded) frames. These frames are arranged into sequences, called Groups of Pictures (GoP), which contains all the information required to decode a given video sequence, within a period of time. It is important to highlight that not all video frames are equal or have the same degree of importance based on 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 [Greengrass et al. 2009a]. Since not all video flows are equal, each video sequence in HCMN systems must be analysed and managed individually or in cluster with group of videos according to the characteristics.

The evolution of the Internet over the past decades aimed to provide network-centric traffic differentiation, resource reservation, and Quality of Service (QoS) support solutions for multimedia distribution applications [Kritikos et al. 2013, Luo and Shyu 2011]. However, QoS proposals only indicate the impact of network performance, such as bandwidth, packet loss rate, and packet delay rate, on the delivery of multimedia content. In this way, QoS schemes alone are not enough to assess, control, and improve the quality level of multimedia applications, because they fail to capture subjective aspects of multimedia content related to human's perceptions. Hence, the distribution of video flows over wireless networking environment in HCMN systems with quality level support, especially in mobile systems, it is not a trivial task and cannot be assure by traditional network or packet-based QoS approaches, i.e., without human and video-awareness.

The perception of video sequences shared and watched by humans must be evaluated in terms of Quality of Experience (QoE) [Lindeberg et al. 2011], which directly measure the video quality according to subjective aspect of the users on watching a given video flow. Thus, many researchers, standardization bodies, and also industries have been studying QoE assessment and management approaches, where the human's experience can be measured and integrated into networking components to improve the overall performance of HCMN systems [Adzic et al. 2011]. This is due to during the transmission of real-time video flows over wireless network, the video quality can be affected by several factors, such as the network infrastructure (i.e., delays and packet loss), the requirements of the human visual system (i.e., regions of interest and environment luminosity), the video characteristics (i.e., video encoder and frame dependency), as well as the user's experience, preference, and interest [Greengrass et al. 2009a]. Hence, QoE and networking schemes must be extended with context-awareness, content adaptation, and human-based models to provide a better application, service, and resource utilization for HCMN systems.

Understanding and modelling human's experiences, including psycho-visual and (re)presentation of user's experience modelling, and video characteristics into wireless systems, decision-making engines, or networking protocols are still open issues. Moreover, it improves the distribution of live multimedia applications with QoE support in HCMN scenarios. Hence, human-centric approaches must be developed to both adapt the videos to different network conditions and human's experiences, as well as to extend networking protocols and services to delivery video sequences with QoE assurance, while optimizing the usage of network resources.

Recent research topics highlighted the importance of content and information for new wired and wireless network services and applications, namely Content and Information Centric Networks (CCN and ICN) [Kulinsk et al. 2013]. However, both CCN and ICN do not enable real-time multimedia adaptation to cope with different network's, device's, and user's requirements. For instance, in congestion periods in a wireless network, the queue scheduling police must decide about which packets should be dropped according to the human visual system, in order to do not affect (or minimize) the user's perception on watching a given video flow. Another example is the use of transcoding services in the network to deliver adapted CCN-based videos to different users. Hence, HCMN has been seeing as a key paradigm to improve the usage of network resources, while also ensuring a better human's experience in consuming or sharing video applications, even in CCN/ICN environments [Lu et al. 2011, Cerqueira et al. 2014].

The vision of HCMN places the users in the centre of multimedia content services, where the video delivery process must be accomplished and optimized according to the human's experience, video characteristics, environmental conditions, display capabilities, and context-awareness as presented in Figure 3.1. In this way, HCMN improves the dissemination of video flows over fixed and wireless CCN/ICN or even in the traditional host-to-host Internet communication. Hence, the dissemination of multimedia content must be optimized by taking into accounting the human's experience, the context-awareness, the user's preference or interest, the human's visual system requirements, the network conditions, as well as the multimedia content characteristics [Wang et al. 2013a].

Figure 3.1. General View of a Human-centric Multimedia Networking Environment

In HCMN scenarios, content creators and network providers should be able to adapt, share, and deliver video flows according to different (cloud) network conditions, devices capabilities, environmental characteristics, and human-centric models or QoE metrics. In this way, several issues need to be studied and understood in order to propose new mechanisms, as well as to extend the exiting ones by taking into consideration the characteristics and requirements of HCMN systems. Several services can be improved with HCMN capabilities, namely handover, routing, packet redundancy, error correction, pricing, and others. For instance, HCMN must improve the system performance and user's satisfaction by taking context, network, and human-awareness into account to create/transmit/access video content. There are several HCMN parameters that can be applied together or separated in application and networking cross-layer schemes, including user's experience, preferences, interests, human visual system requirements, video motion and complexity levels, frame type and importance, network impairments, and mobile device characteristics, and not only network metrics, such as delay, packet loss or signal strength [Cerqueira et al. 2014, Khan et al. 2010]. A trade-off between performance and accuracy must be considered before extending and implementing all or a set of HCMN features into multimedia systems.

The implementation of HCMN system is a hard task, and also requires high processing, data exchange, and memory. However, advances in cloud computing [Wang et al. 2013b] and the integration of HCMN into cloud ecosystems allows mobile users to have new media experiences, as well as the HCMN-awareness that are not possible from their mobile smartphones or tablets, multimedia player in Vehicular Ad-Hoc Networks (VANETs), or even in laptops at home. Moreover, Software Defined Networks (SDN) will also play an important role in HCMN, where it will be possible to improve the control of network resources and trigger video adaptation procedures on-the-fly, as well as evaluate the proposed solution in productive networking environments.

In this work, we start by introducing the concepts behind HCMN systems. In Section 3.2, we introduced the concepts and of key technologies related to video transmission over wireless networks for HCMN scenarios, as well as the methods for video quality assessment based on the user's experience. In Section 3.3, we introduce the cross-layer parameters that must be considered for creation, transmission, or assessment of video over HCMN systems. Section 3.4 presents the state-of-the-art in video delivery with QoE-awareness, which aims to provide video transmission over wireless networks with HCMN capacity. More specifically, we outline the advantages and disadvantages, research opportunities for extending them, as well as possible applications. The final considerations are discussed in Section 3.5, where we indicate the main challenges, future trends and scenarios for HCMN.

3.2. Technologies Related to HCMNs

The human visual quality perception for real-time videos is highly influenced by the networking delivering process, such as packet losses. Moreover, the multimedia transmission over wireless networks is subject to communication errors due to physical network conditions, such as multipath fading, interference, shadowing, and background noise [Adzic et al. 2011]. In contrast to wired networks, packet loss in wireless networks does not mean network congestion, since often failures are related to physical reasons, which causes communication problems and failures [Lindeberg et al. 2011]. Therefore, multimedia services transmitted over wireless networks can be affected by many factors.

Another important key issue in HCMN is the CODEC, where information about the video characteristics can be used to extend networking services and protocols with video-awareness and also model the human's perception. Hence, image compression algorithms are needed to remove spatial and temporal video redundancies, especially in wireless scenarios. In this way, it reduces the amount of data required to depict a video flow, enabling a better use of network resources [Hanzo et al. 2007].

In this section, we introduce relevant concepts related to multimedia transmissions over wireless networks, such as characteristics and problems of wireless networks and video compression techniques for HCMN systems. We also present the main concepts related to QoE, the differences between QoS and QoE, and metric commonly used to estimate the quality level of video sequences.

3.2.1. Video Distribution in Wireless Networks

In this section, we present some use-case wireless network scenarios for video distribution, without addressing technical details about wireless communications. Multimedia applications and services over wireless mobile networks are becoming an integral part of our personal and professional lives. For instance, users will produce, share, and consume many types of content on smartphones in ubiquitous way, ranging from provider-generated entertainment videos to user-generated disaster or surveillance real-time video flows [Miao et al. 2012]. In this context, many mobile wireless devices will use different wireless technologies to share live video flows in HCMN wireless systems, such as by using Wireless Mesh Network (WMNs) [Benyamina et al. 2012], Mobile Ad-hoc Network (MANET) [Kumar et al. 2010], VANETs [Jarupan and Ekici 2011], Flying Ad Hoc Networks (FANETs) [Bekmezci et al. 2013], Long Term Evolution (LTE) [Araniti et al. 2013], or even heterogeneous networks composed of a set of wireless technologies.

WMNs [Benyamina et al. 2012] are now increasingly deployed, enabling users to share, create, and access video streaming with different characteristics or content, such as video surveillance and football matches scenarios. It is possible to see WMNs in campus, stadiums, squares and buildings, where videos are key applications. WMN does not have a fixed network structure or a defined and immutable path to transmit data, because paths are dynamic and discovered in real-time. Hence, we can assume that every router is responsible for maintaining the information flow between other routers in the neighbourhood. However, video streaming produces a degraded performance in multi-hop WMNSs, due to network/channel impairments, such as packet loss. Understanding and modelling the relationship of network impairments, video characteristics, and human experiences are main requirements for the delivery of visual content with HCMN-awareness.

Multimedia transmissions over MANETs [Kumar et al. 2010] have been attracted considerable attention from both academic and industrial research communities over the last years, due to the growth of new multimedia services under ad-hoc networks. For instance, nowadays, more users use MANET to share video content than wired infrastructures. In addition, in case of a natural disaster, such as earthquake, hurricane, or flooding (e.g., hurricane Sandy in New York in 2012 or the combination of floods, mudslides, and landslides in Rio de Janeiro in 2011), the recovery process demands an efficient and rapid deployment of a communication system due to the fact that the standard telecommunication infrastructure might be damaged. In this use-case scenario, MANETs enable to build a temporary communication network to share video flows of monitored area [Morgenthaler et al. 2012]. Peculiar characteristics of MANETs, such as mobility, dynamic network topology, energy constraints, lack on centralized infrastructure, and variable link capacity, make the delivery of multimedia content with quality from the human's experience over these networks a challenging and non-trivial task. Therefore, new HCMN protocols and services for MANETs must be created to transmit adaptive and QoE-aware video flows according to different MANET conditions.

The distribution of real-time multimedia content over VANETs

[Jarupan and Ekici 2011] becomes a reality, which allows drivers and passengers to have new experiences with on-road videos, ranging from surveillance to entertainment services. For instance, Americans spent around 5.5 billion hours in traffic in 2011, where they could have use these hours watching entertainment or educational videos to make their travels more enjoyable. The content providers also recognize VANETs as a promising market for video-based commercial advertisements. Moreover, multimedia vehicles can be used for capturing and sharing environmental monitoring, surveillance, traffic accidents, and disaster-based video flows. Live streaming video flows provide users and authorities (e.g., firefighters and paramedics) more precise information than simple text messages, and also allow them to determine a suitable action, while reducing human reaction times.

Multi-flow video transmissions over FANETs [Bekmezci et al. 2013] enable a large class of multimedia applications, such as safety & security, natural disaster recovery, environmental monitoring, and others. This kind of network is becoming popular and it is possible to buy low cost UAVs equipped with a video camera, an image encoder, and a radio transceiver in many retail stores. FANETs with HCMN capabilities must be able to adapt to topology changes, and also to recover the quality level of the delivered multiple video flows under dynamic topology situations. The user's experience on watching the live video sequences must also be satisfactory even in scenarios with network congestion, buffer overflow, and packet loss ratio.

The evolution of heterogeneous networking access technologies [Roy et al. 2011], real-time multimedia applications, and protocols created a plethora of new wireless connectivity scenarios featuring an ever-increasing number of devices and multimedia networking entities. This heterogeneous multimedia smart environment requires changes in protocols and services to deal with multiple interfaces systems. Hence, the integration of heterogeneous networks in such scenario, such as IEEE 802.11, IEEE 802.16, and LTE in multi-access and multi-operator systems, is bringing about revolutionary changes in wireless environments by providing new opportunities, introducing better communication channels and raising the possibility of providing HCMN approaches for users of wireless services.

3.2.2. Video Characteristics

Video compression techniques aim to reduce the amount of data required to store digital video images, and use both image compression and motion compensation techniques. This is owing to a video recorded with no compression generates a large file, being hard to manipulate and distribute through wireless network. To improve the video delivery process, video compression generates smaller files, which increases its storage efficiency and also enables it to be distributed through the network [Hanzo et al. 2007]. This is especially needed in network scenarios with limited resources, i.e., bandwidth, energy, and memory space.

The CODECs play an important role in HCMN system. H.264/MPEG-4 is still one of the most popular CODECs, but Google VP9 and MPEG-DASH are gaining a lot of space in multimedia environments. MPEG-DASH uses standard HTTP protocol. It can be deployed using standard web servers and it works with existing Internet infrastructures, including caches, firewalls, and NATs. MPEG DASH supports Scalable Video Coding (SVC) and Multiview Video Coding (MVC). The MPEG-DASH specification defines the Media Presentation Description (MPD) and the segment formats. A benefit of MPEG DASH for HCMN systems is that it provides adequate information to the client for selecting and switching between streams, for example, selecting video between different camera angles, and dynamically switching between different bitrates of the same video camera. The delivery of the MPD and the media-encoding formats containing the segments, as well as the client behavior for fetching, adaptation heuristics, and playing content, are outside of MPEG-DASH's scope. Thus, there are a lot of rooms for improving the delivery of MPEG-DASH in HCMN scenarios.

H.264/MPEG-4 is currently one of the most commonly used formats for the recording, compression, and share video content and we will give more attention to it in the remainder of this section. Figure 3.2 shows a generic example for video coding and decoding in a wireless network. At the sender side, the codec encodes the video provided by either a live source camera or stored video repository before transmitting it. Thus, the encoding process removes redundant information and converts the video to an intermediate format (bitstream) so that it can be distributed through the network. The intermediate format is a set of bits created by the encoder in accordance with well-defined standards, such as Moving Picture Experts Group (MPEG) [Le Gall 1991]. Video coding standards, such as, MPEG, specify the bitstream format and the decoding process for a given video sequence, where each flow starts with a sequence header, followed by a GoP header, and then by one or more coded frames. At the receiver side, the H.264/MPEG-4 CODEC decodes the received data, and converts it from an intermediate format to a video sequence [Gualdi et al. 2008].

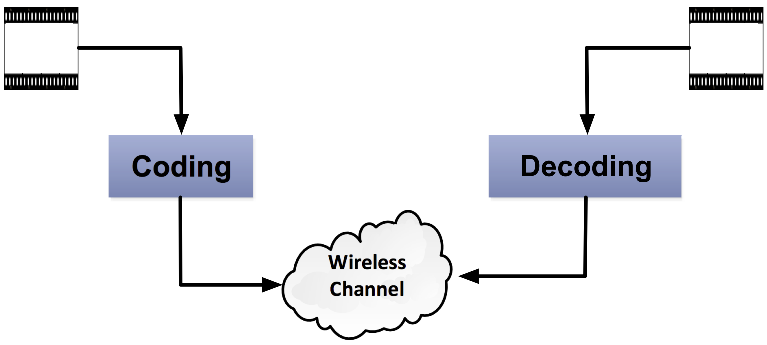


Figure 3.2. Example of video coding and encoding in a wireless network

Both video coding and transmission processes have an impact on the final video quality at the receiver side. Hierarchical video coding schemes, such as MPEG or H.264, convert and compress a video signal into a series of pictures or frames. Change might occurs between one frame and the next, which means that an encoder compress the video significantly by only transmitting the differences [Greengrass et al. 2009a].

3.2.2.1. Spatial or Intra-frame Compression

Intra-frame compression removes redundant information within frames by taking advantage of the fact that pixels within a single frame are related to their neighbours. This process includes signal transform, quantisation, and entropy encoding which is very similar to that of a JPEG still image encoder [Greengrass et al. 2009a]. The intra-frame compression process consists of three stages: computation of the transform coefficients; quantization of the transform coefficients; and conversion of the transform coefficients into pairs after the data has been rearranged in a zigzag scanning order (see Figure 3.3) [Gall 1992].

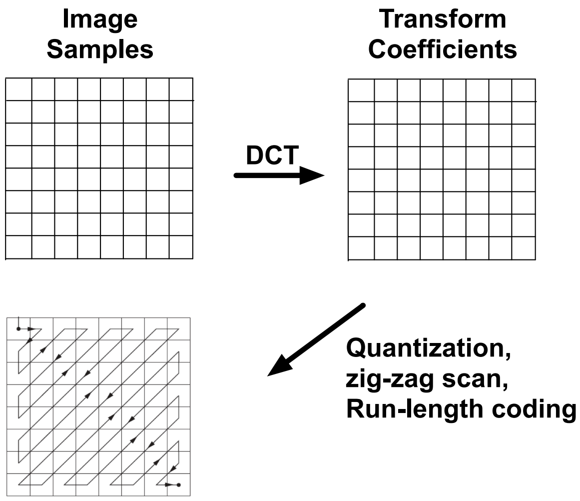


Figure 3.3. Transform coding, quantization and run-length coding

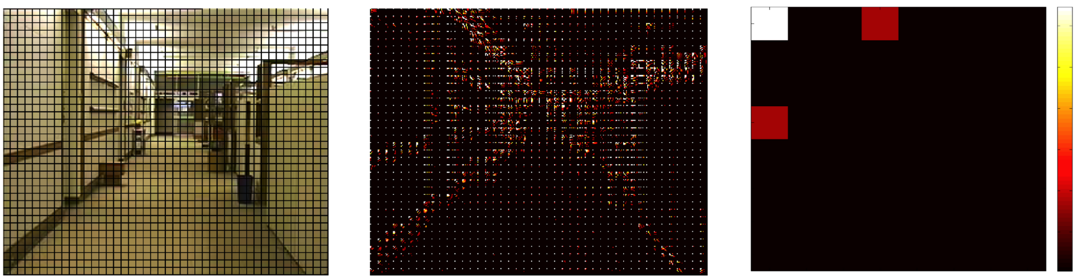
The intra-frame compression uses the Discrete Cosine Transform (DCT) derived from still image compression. DCT extracts signals into a sum of cosine functions that fluctuate at different frequencies. This is because a spatial compression performs frequency analysis in a given frame to find the dominant frequencies, which is carried out by converting frames to the frequency domain by means of transform techniques [Watkinson 2004]. The DCT result is pre-multiplied by the quantisation scale code and divided by the element-wise quantisation matrix. The processed result usually generates a matrix with values primarily in the upper left-hand corner. The zigzag ordering groups all non-zero values. The intra-frame compression greatly reduces the size of the data storage.

Figure 3.4 exemplifies the complexity level for different video sequences downloaded from a well-known video source library [Library 2014]. Figures 3.4(a) and 3.4(d) show a given frame from the Flower and Hall video sequences, respectively. Furthermore, macroblocks can be observed in both frames, which are the basic unit for video frame compression. It divides each frame into small blocks for further handling, and uses the YUV system, i.e., a common used colour space that takes account of human's perceptions when encoding an image or a video. It was initially used in the H.261 standard, and nowadays it is the basis for all previous and current video-coding standards.

The size of the macroblock is variable, but the standard size comprises an array of 8x8 pixels. The DCT transform is applied to each macroblock, and produces a coefficient for each pixel macroblock. In this way, each macroblock has 64 pixels (8x8), i.e., the DCT transform produces an 8x8 matrix containing 64 coefficients. To illustrate this process, Figures 3.4(b) and 3.4(e) show macroblocks obtained after the DCT transform has been applied in Figures 3.4(a) and 3.4(d), respectively. In seeking to give a better explanation of the results of the DCT transform, we used coefficient values in black and white colour scale, where black means lower coefficient values, while white means higher coefficient values. It can be seen that the Hall frame has more macroblocks with a colour closer to white, which explains the higher number of coefficient values in Figure 3.4(b) than in Figure 3.4(e).

Figures 3.4(c) and 3.4(f) show the horizontal and vertical coefficients of two specific macroblocks from Figures 3.4(b) and 3.4(e), respectively. In the matrix containing the coefficients of a given macroblock, the frequency increases from the left to the right and from the top to the bottom. On the basis of these observations, the MPEG saves the coefficients in a vector with an ascending order of frequency.

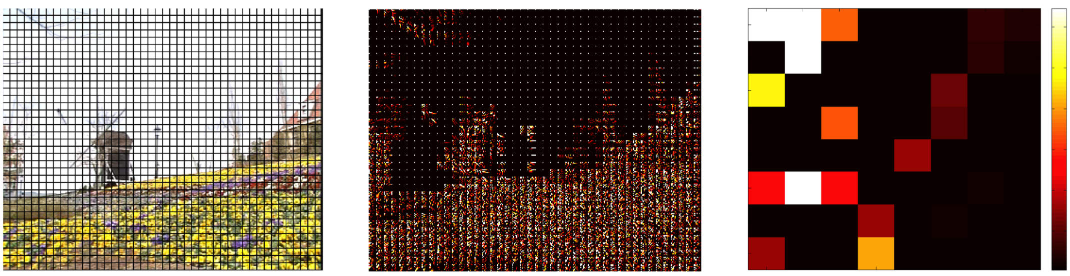
Hence, coefficients with higher frequencies and values closer to zero cannot be transmitted without affecting the level of video quality. In this way, spatial compression reduces the number of bits required to depict a given video frame. However, the spatial compression rates provided by the DCT transform changes, since it depends on the frequency of the image. For example, the Hall frame in Figure 3.4(b) has lower coefficient value for higher frequencies than the coefficients of the Flower frame in Figure 3.4(e). Hence, by calculating the number of DCT coefficients, it is possible to infer the spatial compression rate for a given video.



(a) Hall Video Frame with Macroblocks Division

(b) DCT Coefficients for Each Macroblocks for the Hall Frame

(c) A Single Macroblocks of the Hall Video Frame



(d) Flower Video Frame with Macroblocks Division

(e) DCT Coefficients for Each Macroblocks for the Flower Frame

(f) A Single Macroblocks of the Flower Video Frame

Figure 3.4. Complexity Level for Different Video Frames

3.2.2.2. Temporal or Inter-frame Compression

Temporal or inter-frame compression tries to remove redundancies existing in consecutive frames, obtaining a high compression ratio. This creates a frame based on the previous frame, by eliminating the common parts of the frames. Motion causes the differences between different video frames. Thus, the video frame size reduces by removing the unnecessary parts in the related video frames, and also by only encoding the motion parts. In this way, it is possible to transmit only the differences between the frames.

For instance, it is possible to obtain the frame-difference, as shown in Figure 3.5(c), by extracting the motion difference from the News frames 1 and 2 [Library 2014], i.e., denoted as frame #2 - frame #1. The black portion means the common parts in both frames, and the other parts mean the variation between them. Thus, it is possible to reconstruct the frame #2 based on the frame #1 and the frame-difference.

(a) Frame #1 from News Video Sequence

(b) Frame #2 from News Video Sequence


(c) Motion Difference for Frame #1 and #2 from News Video Sequence


(d) Frame #1 from Football Video Sequence


(e) Frame #2 from Football Video Sequence

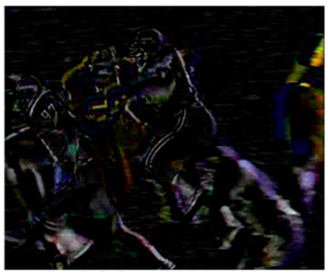

(f) Motion Difference for Frame #1 and #2 from Football Video Sequence

Figure 3.5. Motion Level for Different Video Frames

Figure 3.5(f) shows the frame-difference obtained by extracting the motion difference from Figures 3.5(d) and 3.5(e). First of all, it is possible to observe that Figure 3.5(f) has fewer parts in black than Figure 3.5(c), suggesting that this video has a higher level of motion. Second, the rate of time compression is lower for videos with a high motion level. Third, with motion compensation, only the filtered video frame is stored instead of the original frame, which reduces the video size, since the filtered video frame contains less information. To decode a given video, the motion vector search algorithm must match the motion part in the previous reference video frame to decode the current video frame. The algorithm for locating the motion part is the key part of video coding.

3.2.2.3. Group of Pictures (GoP)

A given video sequence contains a series of video frames, and each frame includes information about the whole picture. MPEG standard employs a hierarchical structure composed of 3 types of frames, namely, I-, P-, and B-Frames [Aguiar et al. 2012]. The macroblock from I-, P-, and B-frames uses spatial compression. On the other hand, temporal compression is only applied to the macroblocks from the P- and B-frames. The macroblocks of the P-frames use as reference the macroblocks from the previous I- or P-frames, as shown in Figure 3.6. The macroblocks of the B-frames use as reference the macroblocks of previous or future I- or P-frames [Serral-Gracià et al. 2010].

These frames are arranged into sequences called Group of Pictures (GoP). A GoP contains all the information required to decode a given video sequence, within a period of time, and can be used for HMCN assessment and optimization procedure. An important factor of MPEG encoding is the GoP size, which indicates the frequency of I-frames in a given video. Each GoP includes an I-frame and all the subsequent P- and B-frames leading up to the next I-frame, as shown in Figure 3.6. For example, a GoP length of 10 frames means a GoP that starts with an I-frame, followed by a sequence of 9 P- or B-frames. For each GoP, M_{GoP} represents the distance between successive P-frames, and N_{GoP} defines the distance between adjacent I-frames. As a result, this structure is flexible, and the frame types and their locations within the GoP can be adjusted to the encoding type. Hence, the MPEG standard provides a hierarchically GoP structured, and some frames are more important than others. In this way, a packet loss has different impact depending on the user's perspective.

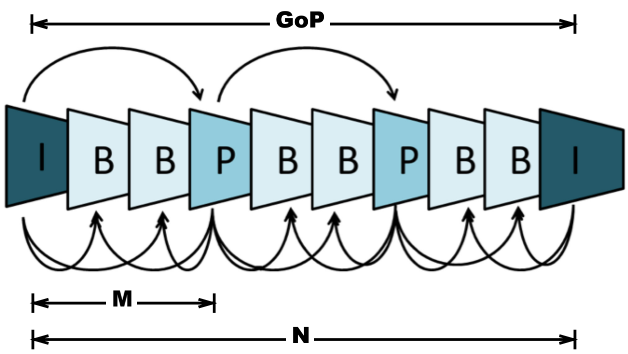


Figure 3.6. Group of Pictures Structure

The frequency of the I-frames in the compressed video defines the size and quality of a video stream. The video bit rate can be reduced by decreasing the frequency of the I-frames when encoding a given video, also degrades the video quality. On the other hand, more I-frames should be added during the encoding process, whenever a higher video quality level is required, and this result in a video with higher video bit rates.

The I-frame contains complete information for a specific video picture, and it is coded without any reference to other frames. In addition, it is known as the key frame for a GoP, since it provides the reference point for decoding a received video stream, i.e., it serves as a reference for all the other frames. It might use spatial compression, and not temporal compression. By removing spatial redundancy, the size of the encoded frame is reduced, and predictions can be used at the decoder to reconstruct the frame. The size of the I-frame is usually higher than the P- or B-frames, since it is the least compressible.

P-frames predict the frame that has to be coded from the previous I-frame or P-frame by using temporal compression, i.e., unlike the preceding I- or P-Frame, P-frames contain the changes of the actual frame. P-frames provide a higher rate of compression than I-Frames, typically 20 – 70 % the size of an I-frame. Finally, the B-frame uses both the previous and the next I-frame or P-frame as their reference points for motion compensation. B-frames provide further compression, typically 5 – 40 % the size of an associated I-frame [Mu et al. 2012].

As a result of the hierarchical structure of MPEG standard, packet losses might affect the level of the video quality in different ways, depending on the lost information [Immich et al. 2013]. More specifically, the loss of an I-frame affects the other B- or P-frames within the same GoP. Thus, errors propagate in other frames until a new I-frame reaches the receiver. In the case of the loss of a P-frame, the error propagates in the remaining P- and B-frames in a GoP. In addition, P-frames that appear earlier in the GoP cause impairments over a longer period, since the subsequent frames are directly or indirectly impaired until the decoder receives the next I-frame. Finally, in the case of B-frame losses, the error does not propagate, since the B-frames are not used as a reference-point for the other frames [Greengrass et al. 2009b]. It is important to notice 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 or to perform HCMN optimization procedures, such as human/QoE-aware packet redundancy schemes.

3.2.3. Video Quality Assessment

Solutions involving multimedia transmissions and HCMN-based must evaluate the video content from the user's perspective and not only from the network's perspective. In this context, over the last decade the focus has shifted away from pure network point-of-view assessment, i.e., QoS metrics, to a more human-centric approach, i.e., QoE metrics and user-awareness, since QoE schemes overcome the limitations related to the human visual system. This is due to QoS schemes alone are not enough to assess the quality level of multimedia applications based on user's experience, because they fail to capture subjective aspects of video content related to human's experience and subjectivity [Mu et al. 2012, Raake and Möller 2011]. In addition, networking services and protocols can the extended and enhanced with QoE features to improve the video delivery process in HCMN environments. It is important to highlight that QoE assessment and management operations are widely dependent on subjective aspects related with human's perception, as well as, user's location, video characteristics, type of CODEC, screen size, and context. For instance, video sequences with different complexities, motions, and frame rates will produce different QoE results even under the same networking scenarios.

QoE metrics have been classified based on a set of manners in the literature

[Fajardo et al. 2011], such as based on metric output type or on the amount of the required reference information. However, in this work, QoE assessment approaches will be defined as objective, subjective, and hybrid as proposed in standardization bodies, working groups, and many works [Moorthy et al. 2010, Wang et al. 2004a]. Objective approaches estimate or predict the video quality level by means of mathematical models or signal processing algorithms. The main objective metrics are the following: Peak Signal to Noise Ratio (PSNR) [Ma et al. 2011], Mean Squared Error (MSE), Structural Similarity Index Metric (SSIM) [Wang et al. 2004b], and Video Quality Metric (VQM) [Pinson and Wolf 2004]. Objective QoE metrics exploit image signal processing techniques to assess the video quality level, but they are difficult to implement in real-time and, usually show a poor performance compared to the human's experience. Moreover, objective metrics fail to capture all the details that might affect the user's experience. This problem is addressed by carrying out subjective QoE evaluations, where Mean Opinion Score (MOS) [ITU-R 2002] is one of the most widely used approaches for subjective video quality evaluation.

SSIM metric improves the performance of the traditional PSNR and MSE metrics. This is because both PSNR and MSE metrics do not correlate well with the subjective perceptions of humans. Therefore, SSIM is a metric that involves frame-to-frame measuring of three components, namely, luminance, contrast, and structural similarity. It also measures the structural distortion of the video, and seeks to obtain a better correlation with the user's subjective impression. Hence, it combines these components into a single value, called index. The SSIM index is a decimal value between 0 and 1, where 0 means no correlation with the original image (low video quality level), and 1 means exactly the same image (high video quality level).

The VQM method defines a set of computational models, which also have been shown to be superior performance than traditional PSNR and MSE metrics. VQM uses the same features of the human's eye to perceive the video quality, including colour and block distortion, as well as blurring and global noise. More specifically, this model employs a linear combination with seven parameters. Four extracted from spatial gradients, two obtained from a chrominance vector, and the last derived from absolute temporal and contrast details. VQM values closest to 0 correspond to the best possible video quality level, i.e., exactly the same image compared to original video. The MSU Video Quality Measurement Tool (VQMT) [Vatolin et al. 2014] is a well-known and suitable tool to measure the SSIM and VQM values for each transmitted video flows.

Moreover, subjective evaluation captures all the details that might affect the human's experience. In this context, MOS is one of the most frequently used metrics for subjective evaluation, and is recommended by the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T). MOS requires human observers rating the overall video quality level in accordance with a predefined scale. The MOS evaluation can be done by following the Single Stimulus (SS) or Double Stimulus (DS) methods defined by the ITU-R BT.500-11 recommendations [ITU-R 2002].

When the SS approach is used, the human observers only watch the video once, and then give a score. The choice of a SS paradigm fits well to a large number of emerging wireless multimedia applications [Seshadrinathan et al. 2010]. When the DS method is applied into the MOS evaluation system, viewers watch an unimpaired reference video, and then they will watch same video impaired. Afterwards, he/she rates the second video using an impairment scale. For both approaches, in general, the MOS scale goes from 1 to 5, where 5 is the best possible score, as shown in Table 3.1.

Table 3.1. MOS Scale According to ITU-R BT.500-11 Recommendations [ITU-R 2002]

MOS	Quality	Impairment
1	Bad	Very annoying
2	Poor	Annoying
3	Fair	Slightly annoying
4	Good	Perceptible, but not annoying
5	Excellent	Imperceptible

Hybrid QoE video quality assessment approaches have been proposed to measure the quality level of videos in real-time by taking into consideration the benefits of both objective and subjective methods. Parametric hybrid schemes predict the perceived video quality level based on information about the IP and video codec headers (bitstream), such as frame type and packet loss rate (without decoding the video flow), where a machine learning technique can be used to map network impairments and video characteristics into a predicted human MOS. An example of parametric approaches have been developed by ITU-T, namely P.NAMS, P.NBAMS, and G.OMVAS.

The P.NAMS is a non-intrusive parametric model for video streaming performance assessment, where it performs procedures based on packet-header information, e.g., from IP through MPEG2-TS. P.NBAMS uses also information about the bitstream for non-intrusive video quality measurement, while P.NBAMS is allowed to use the payload information, i.e., coded bitstream. Although, ITU recommendations include clear descriptions of potential assessment models, few of them have been fully implemented and evaluated with wireless multimedia networking scenarios. The Pseudo Subjective Quality Assessment (PSQA) [Ghareeb and Viho 2010] and Hybrid Quality of Experience Prediction (HyQoE) [Aguiar et al. 2012] are examples of hybrid QoE video prediction approaches available in the literature.

For example, HyQoE uses information about the video's characteristics, network's impairments, human's experience, and Multiple Artificial Neural Networks (MANN) to predict ongoing video flows into MOS. HyQoE collects the video characteristics and network impairments by using a deep packet inspector module, where video coding standards, e.g., MPEG, specify the bitstream format and the decoding process in a video sequence. More specifically, 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. Hence, HyQoE uses a correlation between DCT coefficients, motion vectors, and frame size to define the level of spatial and temporal video's characteristics. HyQoE can be used together with HCMN optimization and management schemes to improve the usage of network resources, as well as system performance in key networking areas, such as pricing, routing, network coding, or mobility.

The implementation of real-time QoE video quality assessment schemes in multimedia network domains is not simple task, and also requires additional studies and experiments. In addition, the interaction between different areas, such as informatics, engineering, visual arts, medicine, and psychology, is crucial to understand the human visual system, behaviour, feeling, and psychological factors. A common sense is that Internet clouds together with mobile clouds will make the implementation of human-centric mobile systems possible, which enrich them with more advanced, cooperative, and powerful features. The development of QoE assessment models by means of crowdsourcing [Gardlo et al. 2014] can also increase the system accuracy and reduce the time required to train human-centric video quality assessment schemes.

3.3. Cross-Layer Parameters for HCMN Systems

The International Organisation for Standardisation (ISO) developed the Open System Interconnection (OSI) model composed of seven conceptual network layers to define the specifications for data transmissions [Zimmermann 1980]. The OSI model provides a universal standard rules for different manufacturers, and also ensures connection compatibility among different software and devices. In this way, we can study the data transmission over networks at different layers, where each layer has specifications to make sure that the data can be smoothly delivered from one device to another device. Each layer is only able to communicate and provide information to the adjacent layers, i.e., based on OSI or Internet (TCP/IP) model, non-adjacent layers are not allowed to exchange information. However, it is an extremely challenging task to meet the end-to-end requirements for multimedia disseminations without interaction between non-adjacent protocol layers [Setton et al. 2005, Srivastava and Motani 2005]. In HCMN environments, it is also important a cross-layer scheme with the humans/end-users, where it is mandatory to collect information about the human's experience to be applied in HCMN assessment and management operations.

The cross-layer design was proposed to enrich services, applications, and protocols with information from more than one layer. It can be defined as the violation of the referenced communication layer architecture by allowing direct communication or sharing of information with non-adjacent layers. In this way, cross-layer optimizations allow flexibility to both have access and to set the information in non-adjacent layers [Foukalas et al. 2008]. This is especially important to assess the video quality level based on user's perception, since those information enables to understand and model the human's experience [Andreopoulos et al. 2006]. In addition, services and protocols can be extended with information about the video application characteristics, such as GoP length, as well as video motion and complexity level. In this section, we describe the key metrics from multiple layers to take into consideration for the design of cross-layer mechanisms for HCMN scenarios.

3.3.1. HCMN Requirements

In HCMN environments, content creators and network providers should be able to adapt, share, and deliver video content in scenarios with different network conditions, devices capabilities, environmental characteristics, and human-centric models. For instance, decision-making process must improve the HCMN system performance, as well as the user's satisfaction by taking into account cross-layer HCMN metrics to create, transmit and access a video flow. The cross-layer context, network, and human-awareness metrics include user's experience, preferences, interests, human visual system requirements, video motion and complexity levels, frame type and importance, network impairments, and mobile device characteristics.

Figure 3.7 depicts a HCMN environment composed of three modules, namely Personalized Devices, Personalized Interfaces, and Context-awareness Engine. Personalized devices, e.g., smartphones, tablets, notebooks, and vehicles (including drones), do not rest only on the computing resources, but on the sensors that they carry, including cameras, light sensors, and location schemes. For instance, video cameras and sensors add important functions to HCMN scenarios, where users (or machines) cooperate with each other to improve surveillance and monitoring missions. Moreover, devices have different display sizes and resolutions, as well as memory and processing capabilities. Thus, all of this information should be taken into accounting for HCMN assessment and networking optimization procedures.

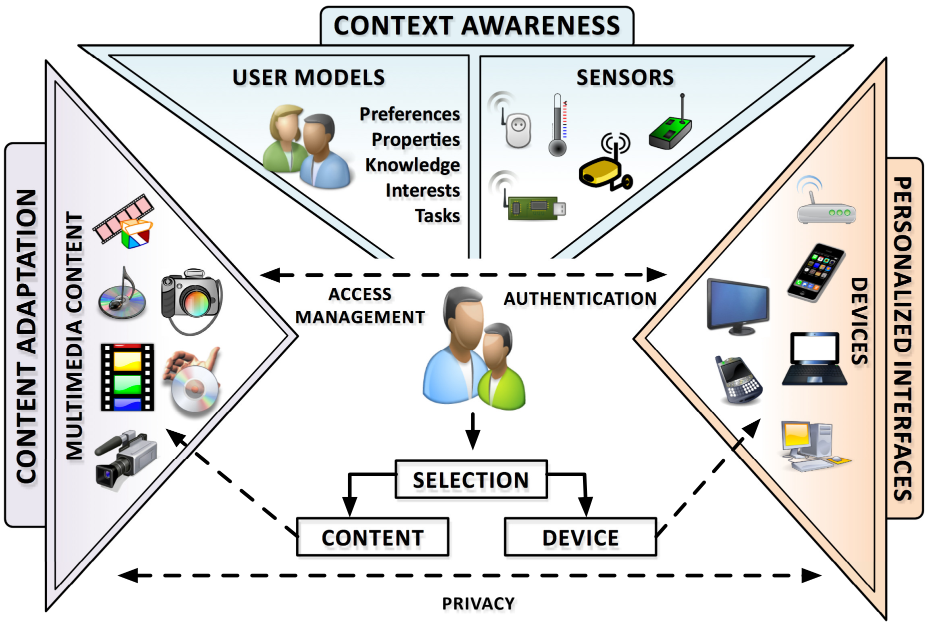


Figure 3.7. Human-centric Multimedia Networking Scenario

Human-centric approaches must also be enhanced with context-aware information. In this context, advanced models describe the human experiences and preferences in consuming real-time video flows, allowing the systems to dynamic adapt the content or resources according to the human needs, properties of the human visual system, and interests [Kiani et al. 2013, La and Kim 2010]. Hence, understanding, modelling, as well as considering user's behaviours, experiences, feelings, and psychological aspects when consuming and sharing real-time video flows are key issues for HCMNs.

Notice that the human-based information for HCMN environments can be collected from social networks. For instance, the user's preferences for a set of videos or even its location can be extracted from his/her Facebook.

3.3.2. Cross-Layer Parameters

New mechanisms must be proposed or extended with the aid of cross-layer parameter in order to improve system performance, user's satisfaction, and also to provide multimedia dissemination with a minimum quality level based on user's point-of-view. Among the services that can be improved with HCMN parameters are video quality assessment, handover, routing, redundancy, error correction, and pricing. In the following, we introduce some cross-layer HCMN approaches that must be considered for decision-making in HCMN system.

3.3.2.1. Network-awareness

Humans expect to watch videos smoothly and at a certain quality level, no matter what changes occur in the networking environment. However, they often experience inconsistent playback, resulting from fluctuation of network quality, especially in case of mobile devices, which have limited bandwidth and hardware resources. As the number of network users increase, bandwidth insufficiency might occur, and thus the multimedia services will be significantly affected. In contrast to general services that accept a certain packet loss rate, multimedia services require low packet loss rate, correctness, sequence order, and also real-time packet transmissions. Moreover, users often view live video flows that freeze intermittent or even failure to operate. Therefore, how to execute smooth playback with limited bandwidth and the different hardware specifications of mobile streaming is an interesting and hard challenge [Lai et al. 2013].

In this context, information about the current and future network conditions, including packet loss and delay rates, quality of the wireless links, congestion levels, available buffers, and system density (sparse or dense) are essential to improve decision-making in HCMN scenarios. The decision-making includes data link to application layers, including networking algorithms, protocols, and mechanisms.

3.3.2.2. QoE-awareness

The evolution of the Internet over the past decades provided network-centric traffic differentiation, resource reservation, and also QoS support for multimedia distribution. However, QoS proposals only indicate the impact of video transmission based on network performance, such as bandwidth, packet loss rate, packet delay rate, and others. Hence, QoS schemes alone are not enough to assess and control the quality level of multimedia applications, because they fail to capture all the subjective aspects from multimedia content that affect the human's perceptions [Mu et al. 2012].

With this goal in mind, several researchers, standardization bodies, as well as industries have been studying QoE optimizations and assessments, where the user's experience can be measured and integrated into networking components for decision-making to improve the overall performance of HCMN systems, as well as to provide QoE-aware video transmissions. In this context, QoE assessment focus on optimizations in the end-to-end HCMN systems based the user's experience.

3.3.2.3. Video-awareness

Video-related information, such as spatial (edges and colours) and temporal (movement speed and direction) characteristics, frame types (I-, P-, and B-frames), intra- and inter-frame dependence, CODECs, GoP lengths, bit rates, resolutions, and region of interests, must be considered for optimizing HCMN approaches. This is due to multimedia content characteristics have great impact on the performance of HCMN systems, such as presented in Section 3.2.2.

For instance, I-, P- and B-frames compose a compressed video, which have different priorities based on user's experience. The loss of high priority frames causes severe video distortion, since some received packets cannot be decoded at the receiver side. It also wastes scarce network resources, such as bandwidth and energy. The energy cost is due to intermediate nodes spend energy to forward a packet, which is not useful to reconstruct the video received at the destination node [Costa et al. 2013]. Moreover, 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, the impact of a packet lost can be annoying or not for the end user [Aguiar et al. 2012].

In this context, video-related information can be collected from deep packet inspection schemes, off-line signal processing analysis, or described in a Media Presentation Description (MPD), such as proposed in MPEG DASH standard. These information are extremely important for video assessment, control, as well as for collaborative HCMN scenarios.

3.3.2.4. Context-awareness

Multimedia applications and services must be aware of the environment/user's contexts to automatically adapt to environment/user changes, known as context-awareness. By context, we refer to any information that can be used to characterize the situation of an entity, where an entity can be a person, place, or physical or computational object [Dey and Abowd 2000]. For example, contexts may include person (name, role, etc.), location contexts (coordinate, temperature, etc.), user's preferences and activity, computational entity contexts (device, network, application, etc.) and activity contexts (scheduled activities, etc.) [Gu et al. 2005]. Hence, context-awareness relies on different user information in order to provide content, resource, and service relevant to his current situation in a transparent way.

Indeed, multimedia services and applications are very sensitive to diverse context information related to different context entities. More specifically, live video services, more than other services, need to benefit from a context-aware framework, enabling to dynamic service personalization, as well as content adaptation to achieve a better QoE in a HCMN system [Chellouche and Négru 2012]. The main difference between conventional and context-aware methods lies in the process performed by service providers after the services gets invoked. For one hand, without context information, a searched service is just bound to the consumer. On the other hand, the context information is accompanied with the service invocation, and analysed to determine the current situation of the consumer. Afterwards, the most appropriate service is selected at the stage of adapting services, and the service is invoked. While a service is running, its associated context can change, and thus the service needs to be dynamically adapted. In summary, the context information provides a key clue to adapt service dynamically in terms of service personalization [La and Kim 2010] and according to HCMN features.

In this context, human-based data can be collected from social networks, social interactions, and/or any context management systems, such as proposed by [Kiani et al. 2013, La and Kim 2010, Rahman et al. 2012]. Moreover, users have devices (smartphones, tablets, notebooks, and other) with different display sizes and resolutions, as well as memory and processing capabilities. Hence, video flows could also be distributed based on the available resources in mobile devices to improve the usage of network resources, as well as increase the user's satisfaction.

3.4. State-of-the-art Analysis on HCMN

In this section, we introduce the well-known HCMN proposals, which improve the video distribution over wireless environments with QoE assurance. Moreover, we present the main tends about the usage of HCMN approaches in networking scenarios, focusing on the user's perspective.

3.4.1. State-of-the-art

In HCMN environments, after understanding and modelling human-centric schemes, video characteristics, and context-awareness modules, decision-making engines, networking services and protocols can be extended or created to improve the transmission of live video content following a HCMN approach. Some existing works in the literature that include one or many HCMN features, including video quality prediction, video/QoE-aware Forward Error Correction (FEC), packet redundancy schemes, routing, handover, and networking adaptive services.

3.4.1.1. Video Quality Level Assessment

Modelling the relationship of network impairments, video characteristics, and human's experiences by using wireless quality level assessment schemes are key requirements to delivery visual content in multimedia mobile networks, such as football matches and other live multimedia events. 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.

As discussed in Section 3.2.3, solutions for assessing the video quality level based on the human's perception can be organized as subjective, objective, and hybrid. Subjective schemes are hard to implement in real-time. Objective video quality assessment technologies are categorized into several parametric model types, 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's perception of video flows is influenced by the number of the edges (spatial information ñ complexity level) in the individual frames, and also by the type and direction of movement (temporal information ñ motion level) in a GoP. However, existing solutions have not been implemented and evaluated in multimedia wireless systems, or presented inaccurate results based on the user's experience, because, mainly, they do not consider the video motion and complexity levels in their assessment procedures.

Video quality assessment approaches should try to model the human's experience when they are watching videos with different motion and complexity levels under different network conditions. Thus, it is possible to predict the video quality level with high accuracy and in real-time. A non-intrusive QoE parametric scheme, called PSQA, has been implemented to predict the quality level of video flows [Ghareeb and Viho 2010]. The authors included a Random Neural Network (RNN) model together with its learning algorithms, in order to assess the quality level of video sequences in real-time based on a set of parameters, including frame type, frame rate, and packet loss rate. PSQA was originally proposed to improve the understanding of QoS factors in multimedia engineering without an in-depth understanding of the user's experience.

In addition, a PSQA extension, called MDC-PSQA [Ghareeb and Viho 2011], introduced a hybrid video quality assessment approach for Multiple Description Coding (MDC) videos over multiple overlay paths. The MDC-PSQA models a QoE prediction scheme that takes into account the human's perception for a single video flow and network conditions. It also uses the GoP length and the percentage losses of the I-, P-, and B-frames in a GoP during the video quality prediction process. However, the MDC-PSQA prediction model should be improved to deal with video flows with different motion and complexity levels, such as expected in future HCMN systems. Furthermore, visual quality metrics must be tested in a wide variety of visual contents and distortion types, before meaningful conclusions can be drawn from their performance.

A QoE video quality prediction solution proposed by [Khan et al. 2009] classified different video flows into groups representing different content types. They used a combination of temporal and spatial levels, as well as extraction features, by means of cluster analysis. Based on the content type, the video quality (i.e., 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). The proposed scheme measures the video quality level by applying the PSNR average to all the decoded frames (off-line process), which performs poorly compared to MOS.

Few works analysed the impact of live video distributions with different motion and complexity levels over wireless networks according to the human's perception. Some proposals focused on estimating the in-service QoE for peer-to-peer live video streaming systems using a user-centric and context-aware approach based on Bayesian networks [Mian et al. 2013]. Other approaches are focused on assess the quality level of real-time video streaming, such as the Hybrid Quality of Experience Prediction (HyQoE) [Aguiar et al. 2012] and its extension MultQoE [Aguiar et al. 2014] proposal, which follows the ITU-SG 12 recommendations. MultQoE defines its specific input video, packet, or network parameters, and validates an accuracy parametric video quality estimator solution for multimedia wireless networks.

MultiQoE uses MANN to map video's characteristics, human's perceptions, and network's impairments into a predicted MOS, as results it achieves QoE assessments closer to human's scores. MANN has been successfully used for QoS or QoE assessment schemes, as well as 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.

An instance of MultiQoE can be obtained by following the procedures defined in five main components, as detailed in [Aguiar et al. 2014]. Each of them is designed to complete single or multiple tasks for the modelling of the quality evaluation model. Figure 3.8 illustrate the components of MultiQoE, namely: (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's Experience, and Network Impairments into MOS.

The Source Video Database (Component 1) classifies Internet videos according to their spatial and temporal (motion and complexity) characteristics. Video content characteristics together with the percentage of I-, P-, and B-frames losses in a certain GoP are used by the Component 4 (Measurement Model of Factors Affecting Quality), in order to identify the video motion and complexity levels as well as the impact of the transmissions frames of each video sequence. It is important to note that to improve the system accuracy, each ANN is responsible for videos with a specific GoP length, such as 10, 20, or 30. At the same time, it keeps a distorted video database composed of videos delivered (as expected to be watched by humans) in real/simulated networks.

The Component 2 (Network Transmission) is responsible for transmitting all videos in wireless networks, which experienced different congestion, errors, and impairments levels. It gets information on packet loss and delay of all 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, such as specified in Component 4. In the Component 3 (Subjective Quality Assessment and Distorted Video Database), a set of humans evaluates all distorted videos (following the ITU recommendations) to define and score their MOS. Finally, Component 5 uses a MANN to correlate video's characteristics, human's experience, and network's impairments into a predicted MOS value.

In this context, MultiQoE can be used together with optimization and management networking schemes to improve the usage of network resources, as well as to increase the system performance in key networking areas, such as, FEC, routing, or mobility.

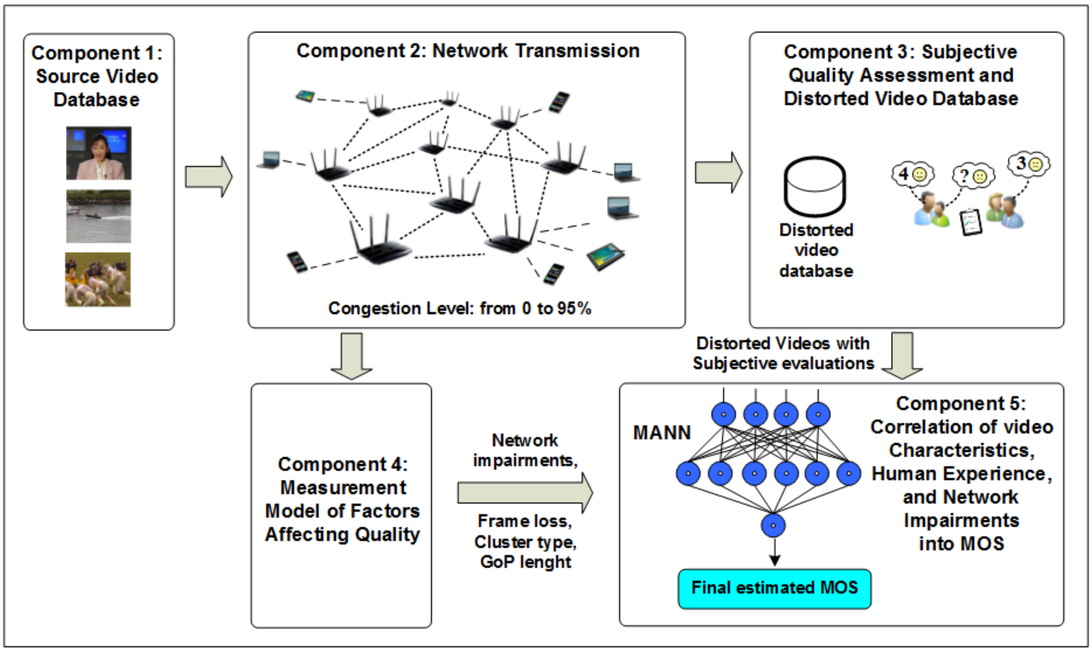


Figure 3.8. MultiQoE components [Aguiar et al. 2014]

3.4.1.2. Packet Redundancy

Node constraints, such as bandwidth, increase the effects of wireless channel errors. Moreover, as soon as video packets are lost or arrive late, there is a significant decline in the resulting video quality based on user experience. In this context, error correction techniques can be applied to provide resilient and robust video transmission over unreliable wireless communication channels. Error control mechanisms that deal with wireless transmission errors in multimedia streaming applications include Automatic Repeat Request (Automatic Repeat Request), Forward Error Correction (Forward Error Correction), and Packet redundancy [Naderi et al. 2012].

Packet-level redundancy mechanism protects video streaming from channel errors without an extra delay. This is because it acts in a complete different way than link-layer mechanisms, and it has been employed due to its suitability for multimedia communications, as well as the nature of the error coding at the application layer. Moreover, ARQ uses the bandwidth in an efficient way compared with the FEC techniques, although ARQ incurs additional latency costs, which can not be tolerated for live video sequences [van der Schaar et al. 2003]. Hence, among the existing error control schemes to handle packet losses in real-time multimedia communication, application-level redundancy mechanisms offer a suitable solution to provide video delivery with quality level assurance, as well as without adding delay and considering end-to-end reverse channel.

Regarding to FEC or packet redundancy mechanisms for wireless networks in HCMN scenarios. [Tsai et al. 2011b] introduced a forward-looking forward error correction (FL-FEC) mechanism to recover lost packets in order to improve video quality level. The redundancy mechanism encodes n packets with (k - n) redundant packets to form a block with the *k* packets at the sender. Then, the mechanism can tolerate the loss of (*k* - *n*) packets in networks and recover the *k* packets from the FEC block at the receiver. The proposed redundancy mechanism recovers not only the lost packet from its block, but also the previous block from the recovered packet. The recovery procedure is repeated until recovering the first block.

Latter, [Tsai et al. 2011a] combined ARQ and FEC mechanisms into an adaptive HARQ mechanism, called the Adaptive Hybrid Error Correction Model (AHECM). AHECM finds appropriate parameters, i.e., maximum retransmission threshold and packet redundancy, to avoid network congestion and also reduce the number of redundant packets by predicting the effective packet loss rate. More specifically, AHECM collects meta-information about the average packet loss rate, the average Round-Trip Time (RTT), and the available bandwidth at the receiver. As soon as the meta-information changes due to channel or network conditions, the AHECM makes the receiver send the meta-information to the sender in order to adjust parameters of the AHECM. Hence, the AHECM at the sender choses the appropriate redundancy for different video frame types when the sender transmits video streaming to the receiver over wireless networks. With information about the average RTT and tolerable end-to-end delay provided by the receiver, the AHECM at the sender finds the maximum retransmission time and retransmits the lost packet to the receiver in time at the premise of reducing the packet redundancy. With information about the average packet loss rate and the available bandwidth provided by the receiver, the AHECM at the sender calculates the appropriate parameter to avoid network congestion and the unnecessary packet redundancy. Meanwhile, when the end-to-end delay requirement can be met, the AHECM only retransmit the necessary number of redundant packets to receiver in comparison with legacy HARQ mechanisms. The AHECM uses a Markov model to accurately predict the effective packet loss rate at the premise of considering the burst bit error condition in wireless networks.

The Adaptive Cross-Layer FEC (ACFEC) mechanism uses a packet-level error correction to determine the amount of packet redundancy to generate in a wireless network scenario [Han et al. 2010]. More specifically, at the MAC layer it verifies when a loss occurs, in order to increase a failure counter. In this way, this information defines the amount of FEC recovery packets. However, one weakness of this proposal based on the HCMN point-of-view is that the video characteristics are not considered, which are known to have a direct influence on the video resilience to packet loss, QoE, and also to reduce the network overhead.

The mechanism, called Cross-Layer Mapping Unequal Error Protection (CLM-UEP), aims to improve the delivery of video streaming over IEEE 802.11e wireless networks by using a video-aware FEC approach [Lin et al. 2012]. CLM-UEP assigns a different level of redundancy to frames with different importance based on the user's point-of-view. It implements an adaptive algorithm for mapping video flows and redundant packets to a suitable Access Category (AC) queue. This procedure also takes into consideration the frame type, packet loss rate, as well as the AC queue occupancy to avoid congestion-induced packet losses. However, this proposal does not consider video flows have different motion and complexity levels, which introduce different impact for a frame loss based on user's experience. For instance, depending of the video characteristics, the first P frame in a GoP can be more important than the second one.

Network coding techniques can be also integrated with QoE-awareness to improve the delivery of multimedia packets over wireless networks in HCMN scenarios. An example of the integration of QoE and network coding is presented in [Pimentel-Nino et al. 2013]. The QoE driven adaptive video with overlapping network-coding proposal aims to enhance the transmission of video flow over satellite links. The authors considered a cross-layer approach to adapt the video quality level (and redundancy) to different network conditions. However, human-based QoE models should be integrated into FEC and redundancy approaches to improve the amount of decoded video packets, while saving network resources.

[Zhao et al. 2012] introduced a QoE-aware FEC mechanism for intrusion detection in multi-tier Wireless Multimedia Sensor Networks to improve the distribution of live video flows. This work creates redundant packets based on impact of the frame loss based on the user's experience. The proposed QoE-aware FEC mechanism targets the reduction of redundant packet transmitted over wireless network, while keeping videos with an acceptable quality level.

As discussed before, an error correction mechanism needed to provide the distribution of video flow with acceptable video quality level, whatever network adversity occurs. Hence, an adjustable FEC-based mechanism must use Unequal Error Protection (UEP) schemes together with a video/human-awareness model, which reduces the amount of redundant information transmitted over the wireless network. With this goal in mind, [Immich et al. 2013] proposed an adaptive cross-layer VIdEo-aWare FEC-based Mechanism with Unequal Error Protection scheme (ViewFEC). It aims to support video transmissions to wireless users, while assuring QoE and optimizing the usage of wireless resources. Based on information about the network conditions and video characteristics (motion and complexity levels), ViewFEC can be optimally configured to send redundant information only to sensitive video sequences from the user's experience.

ViewFEC relies on a set components and stages, such as presented in Figure 3.9 and detailed in [Immich et al. 2013]. At Stage 1, ViewFEC relies on a video classifier to fetch information from the Cluster Analysis Knowledge basE (CAKE), and also the Cross-LAyer inforMation (CLAM) components to identify key video characteristics, including motion and complexity levels, as well as GoP size. The above information identifies the impact of different video flows based on the human's visual system. Afterwards, at Stage 2, further details about the video flow are gathered, namely type and relative position of the frames within its GoP. Finally, at Stage 3, the FEC blocks are built and an UEP redundancy is assigned to each one. The simulation results show that the ViewFEC outperforms non-adaptive Video-aware FEC-based schemes in terms of recovery rate, video quality level, and especially network overhead. There is no need to protect all packets of a frame to obtain a video quality improvement from the user's point-of-view because codecs are resilient to a certain amount of loss, especially at the end of the GoP.

3.4.1.3. Routing

Routing schemes are also important networking services that must be integrated into HCMN approaches. In this way, routes must be selected not only based on traditional

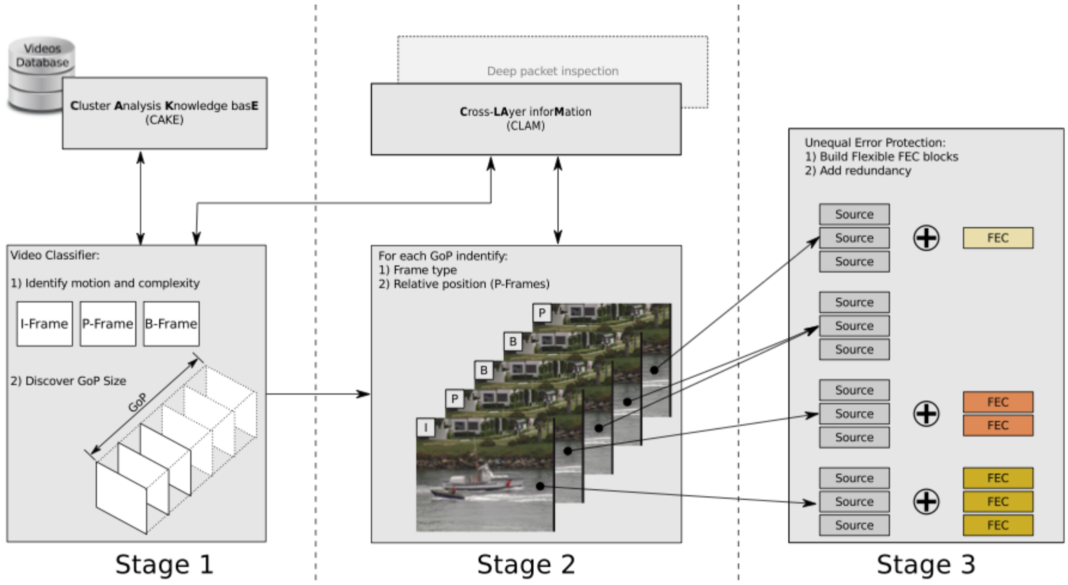


Figure 3.9. ViewFEC stages [Immich et al. 2013]

QoS metrics, but the routing service must define routes for video delivery according to the user's point-of-view. A user QoE-based adaptive routing system for future Internet relies on PSQA to predict the video quality level, and as result it selects the best QoE routes for video packets [Tran et al. 2012b]. The authors proposed a protocol based on user QoE measurement in routing paradigm to construct an adaptive and evolutionary system. An extension of the proposed routing scheme is presented in [Tran et al. 2013]. However, in both works, the results presented a significant performance against other traditional pure QoS routing protocols.

[Tran et al. 2012a] introduced the Dynamic Optimized QoE Adaptive Routing (DOQAR), which aims to improve the user's perception and to optimize the usage of network resources in a wired network across a wireless access network (end-to-end fashion). Experimental results show that this routing protocol gives significant QoE evaluation improvements over traditional routing approaches. Moreover, a hierarchical routing management enhances the transmission of multimedia content with QoE support [Volkert and Mitschele-Thiel 2012]. In the proposed scheme, video streams are controlled and routed by a hierarchical routing management system to select the best paths for multimedia flows. This leads to an improved QoE of received multimedia streams. However, in both solutions, improvements in the video quality assessment and routing processes are still needed.

The routing framework for QoE-based routing in multi-service Wireless Mesh Networks (WMNs) aims to increase the user's experience when watching a video [Matos et al. 2012]. The proposed solution takes into account the heterogeneous requirements of different services delivered over a WMN, such that the overall end-user QoE is maximized under given resource constraints. A double reinforcement learning strategy is used to dynamically compute the most efficient routes to deliver the flows of each service type with QoE support.

The HCMN services and applications rage from entertainment and human digital memory video to user-generated disaster or surveillance real-time video flows. For instance, in case of a natural disaster, such as Hurricane Sandy in New York/USA (2012), flooding in Rio de Janeiro/Brazil (2013) or any other disaster environments, the standard telecommunications infrastructure might be damaged or does not exist anymore. Hence, a group of UAVs equipped with video camera could be used to set up a temporary multimedia Flying Ad-Hoc Network (FANET) with the aims to explore, sense, and also send multimedia data from the hazardous area, enabling humans in the control center to be aware of what is happening in the environment, as well as take action based on rich visual information. For such multimedia FANET scenario, video flows collected by a given UAV must be transmitted by taking into account HCMN features. In this way, it is possible to deliver video flows with QoE assurance to headquarters or computer systems for further processing, analysis, and dissemination, such as provided by the semantic system [Macias et al. 2012], sensor4cities [Lima et al. 2012], i-SCOPE [i SCOPE], and any other platforms.

In this context, [Rosário et al. 2013b] introduced a Cross-layer Link quality and Geographical-aware beacon-less OR protocol (XLinGO) for real-time video delivery in mobile networks. XLinGO can be applied for many FANET multimedia applications, such as safety & security, environmental monitoring, natural disaster recovery, and others. In terms of performance, XLinGO enables an efficient and reliable multiple video flows transmission together with QoE assurance and low overhead over multimedia FANET scenarios, by taking into account metrics from multiple layers.

XLinGO relies on across-layer approach to improve the routing and also packet redundancy decisions according to application-layer (video characteristics and requirements), link layer (link quality), energy, geographical information, and human visual system information. More specifically, the routing service includes forwarding and MAC functionalities, where it assumes a CSMA/CA mechanism, and relies on beaconless Opportunistic Routing (OR) method. The beacon-less OR approach helps the nodes to transmit live video sequence, where nodes forward packets to the destination node based on a distributed hop-by-hop routing decision and without a stable end-to-end route. To create and keep reliable and robust persistent multi-hop routes, XLinGO combines packet delivery ratio, QoE, queue length, link quality, geographical location, and residual energy, as depicted in Figure 3.10.

Moreover, XLinGO introduces a recovery mechanism to deal with route failures, providing a smoother operation in harsh environments and mobile networks. Third, it applies a QoE-aware redundancy scheme to add redundant packets only for important video frames, reducing the network overhead, while maximizing the human's experience and save scarce network resources.

3.4.1.4. Handover

The handover procedures in HCMN systems must be accomplished based on QoE assessment, and this issue has been discussed in many papers available in the literature. Among them, a mobility framework to enhance the QoE through QoS to support high-quality

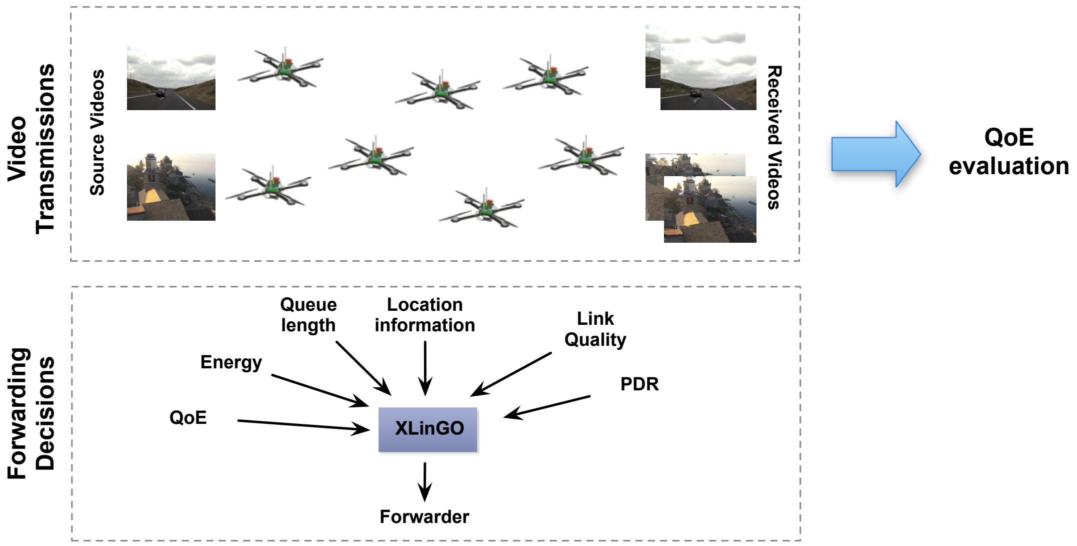


Figure 3.10. XLinGO forwarding decisions

services for on-going multimedia applications [Khorsandroo et al. 2012]. The QoE handover performs its procedures based on a dynamic QoS provisioning system for network mobility. It also uses a quantitative relationship between QoS and QoE to give support for handover decisions. However, the proposed human's perception model should be improved by taking into accounting information about video with different characteristics, and also user's experience.

[Politis et al. 2012] proposed a seamless handoff scheme that incorporates IEEE 802.21 Media Independent Handover framework, and a QoE-driven rate adaptation scheme, for both scalable and single layer coding. A good point in this work is the use of QoE during the network selection process. The inclusion of human's experience during the handover process aims to improve the mobility of video flows between different access points/networks. Moreover, a similar approach is also exploited in [Rosário et al. 2013a].

Existing QoE handover schemes rely on machine learning techniques to model the human's behaviour to enhance the handover process during video transmission. In this context, [Ghahfarokhi and Movahhedinia 2013] proposed a personalized QoE-aware handover decision based on distributed reinforcement learning. The paper presents a personalized human-centric handoff decision method to decide about the time and target of handover based on User Perceived Quality (UPQ) feedbacks. Hence, users watching video sequences can be always best connected to the most suitable access point. However, they used only PSNR evaluation to show the benefits of the proposed solution based the user's experience, where PSRN assessment performs poorly compared to other QoE objective evaluations, as mentioned in Section 3.2.3.

[Varela and Laulajainen 2011] introduced another QoE handover solution, which takes QoE into consideration for handover decisions. The proposed scheme uses the PSQA assessment to predict the quality level of video sequences in different access points/networks. The handover module uses the predicted QoE of on-going videos in candidate networks as an indicator to select the best network for connection. It also balances the trade-off between user's profit and overall network condition by taking into account overall user satisfaction when making decision.

[Taleb and Ksentini 2012] designed an admission control mechanisms that helps mobile operator to cope with handover decision process between macro and femtocell networks. It considers QoE evaluations for a femto multimedia communications, providing QoS/QoE-based admission control mechanisms. By using PSQA, the admission control mechanism predicts the user's experience in both networks and improves the handover decision process for video flows. The proposed solution uses user's feedbacks as input for a QoS/QoE mapper scheme that learns the relation between user's satisfaction and current QoS conditions of a cell (data rate mainly).

[Shehada et al. 2013] proposed a QoE-based mechanism to provide smooth handover for video flows over LTE wireless networks. The proposed solution reserves resources (statically or dynamically) in order to maintain unperceivable quality fluctuation during handover periods for video sequences in LTE mobile networks. Figure 3.11 presents the proposed QoE-based handover and resource reservation approach. All videos are stored at high quality at the Application Server and each user device can access the videos. The Traffic Management module in the core network acts as a downlink resource allocator. The Traffic Engineering component is a rate shaper. Network resource allocation selection is done in a cross-layer fashion, taking into account the utility function of the video watched by the end-user and the wireless channel condition. Thus, it is possible to model the human's experience for different network performance parameters.

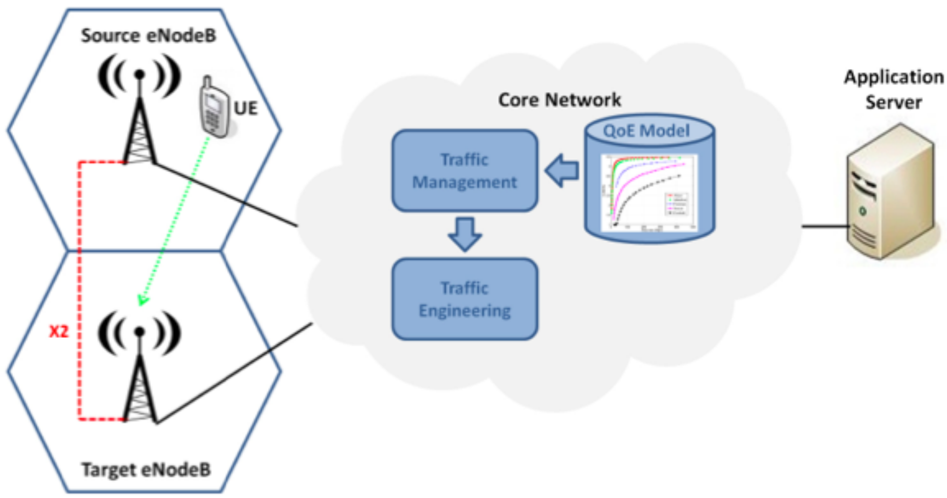


Figure 3.11. QoE-aware multi-cell handover scenario [Shehada et al. 2013]

3.4.1.5. Management

QoE-based network management operations are essential to disseminate multimedia content for humans, while improving the usage of network resources. With this goal in mind, [Mu et al. 2009] introduced a QoE-aware real-time multimedia management, which provides an end-to-end quality control on real-time multimedia applications over heterogeneous. It integrates video QoE assessment together with QoS and QoE-based mapping for adaptation procedures. The main goal of the proposed solution is to extend the network management functions to recognize, plane, and distribute multimedia content with QoE-awareness, while improving the efficiency of network resource utilization.

[Seppanen and Varela 2013] proposed a QoE-driven network management for real-time over-the-top multimedia services. It supports network-level management mechanisms for packet traffic, while using QoE as a performance indicator. The access points implement three main modules for QoE-aware procedures, namely QoE assessment, traffic classification, and traffic management. The QoE assessments component predicts the video quality level based on the PSQA mechanism. The traffic classification classifies the flows into a pre-defined class of services, including premium and normal interactive, streaming, and bulk traffic applications. The traffic management performs QoE-based inter-class scheduling, intra-class scheduling, bandwidth management, and admission control procedures.

[Seppanen et al. 2013] extend traffic control mechanisms with QoE-awareness in order to provide an autonomous QoE-driven network management framework. The proposed approach is able to identify media flows and predict the QoE of video flows based on their application type, subscriber class, current QoE for it and other media flows, and expected QoE after control mechanism operations. After that, the system is also able to perform access control on new flows based on the current quality for existing flows, and the incoming flow's application and subscriber class.

[Taboada et al. 2013] introduced another interesting QoE networking approach, which shows how to incorporate human perceived quality in the design of resource allocation algorithms. A key contribution of this work is related to the modelling of the QoE maximizing problem for scheduling multimedia flows in a shared channel. The proposed QoE-aware scheduling policy can be useful for network providers in order to guarantee adequate satisfaction levels to their customers when watching multimedia content.

Mobile video delivery has become a major challenge for mobile operators. With the demand for multimedia content, wireless access congestion is becoming more frequent, degrading the quality level of mobile video. Therefore, [Fu et al. 2013] introduced a QoE- aware traffic management scheme for scalable video delivery. The proposed optimization function aims to maximize the overall QoE of multiple users by allocating optimal transmission bitrates to users. The authors consider bandwidth conditions to estimate the maximum achievable bitrate of each user. To understand the effect of traffic management on QoE, utility functions are used to understand and model the relation between transmission bitrates and perceptual quality in terms of MOS values.

QoE-based procedures can also be used to improve the delivery of video content in cognitive radio systems. In this context, [Imran et al. 2013] mapped the human's perception in scenarios with different throughput, as well as packet delays and losses. The QoE enhancement in MIMO cognition through the Multibeam Opportunistic Beamforming technique is the main objective of the proposed solution, where statistical optimization is being employed within such scenarios. Thus, it is possible to keep video flows with a better quality level during congestion periods in cognitive radio networks.

The use of multimedia-aware caching approaches aims to improve the delivery of multimedia content to users. For instance, [Zhang et al. 2013] studied the problem on how to cache a set of media files with optimal streaming rates, under HTTP adaptive bit rate streaming over wireless networks. The proposed system implemented an optimization framework to maximize the QoE objective function for a given storage budget. The cache management scheme aims to provide high QoE, while requiring low complexity, which gives guidelines for practical design of HTTP adaptive bitrate streaming services.

Adaptive video streaming is key solution to increase user QoE and maximize connection utilization. The integration of MPEG DASH (Dynamic Adaptive Streaming over HTTP) and OpenFlow/Software Denifed Networks can improve the delivery of video sequence to mobile users and the usage of network resource. Therefore, [Georgopoulos et al. 2013] proposed an OpenFlow-assisted QoE Fairness Framework (QFF), which aims to fairly maximize the QoE of multiple competing clients in a shared network wireless system.

The QFF proposal uses OpenFlow allows vendor agnostic functionality to be implemented for network management and active resource allocation. OpenFlow allows QFF to monitor the status of all the DASH video flows in a network and dynamically allocates network resources to each device, while achieving the maximum user-level fairness. QFF is composed of a set of components, including utility and optimization functions. The utility function aims to model the human's experience based on the mapping the bitrate of a video at a particular resolution and the QoE perceived by the user. The optimization function uses the human's perception models to find the optimum set of bitrates that ensures QoE fairness across all DASH clients in the network.

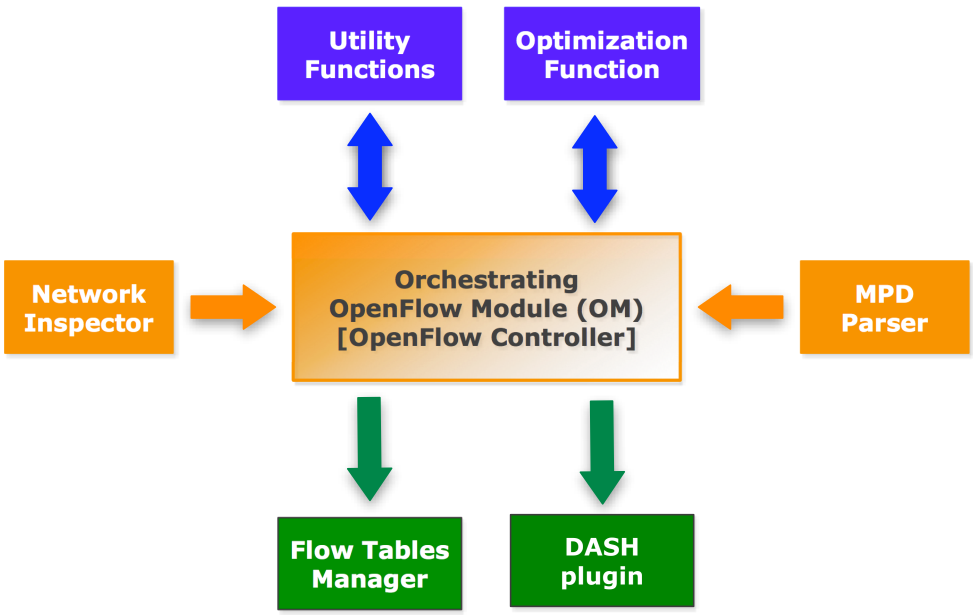


Figure 3.12. OpenFlow-assisted QoE Fairness Framework [Georgopoulos et al. 2013]

The QFF proposal uses OpenFlow allows vendor agnostic functionality to be implemented for network management and active resource allocation. OpenFlow allows QFF to monitor the status of all the DASH video flows in a network and dynamically allocates network resources to each device, while achieving the maximum user-level fairness. QFF is composed of a set of components, including utility and optimization functions. The utility function aims to model the human's experience based on the mapping the bitrate of a video at a particular resolution and the QoE perceived by the user. The optimization function uses the human's perception models to find the optimum set of bitrates that ensures QoE fairness across all DASH clients in the network.

3.5. Final Considerations and Future Directions

In this work, we introduced the HCMN approach for live video distribution in next generation of wireless multimedia networks, which is becoming a reality for users, industry, and academia. HCMN will improve multimedia assessment and transmission procedures and will help multimedia systems to continue expanding their portfolio of applications during the next years with adaptive real-time video schemes. In HCMN scenarios, thousands of users will produce, share, and also consume multimedia services in a ubiquitous manner on smartphones or tablets, even in vehicles, UAVs, and other mobile devices. The multimedia services and applications rage from entertainment and human digital memory video to user-generated disaster or surveillance real-time video flows. For instance, CISCO projected that over 10 billion devices will be in used until 2016, and 71% of all mobile data traffic is expected to be videos by that time as well. Moreover, the multimedia transmission will represent up to 90% of the global IP data traffic in a few years. In addition, the real-time multimedia transmissions to thousands of people over wireless network will take place in the coming big events hosted in Brazil, such as in the soccer World cup and the Olympic Games in Brazil in 2014 and 2016, respectively.

The vision of HCMN places the human's experience in the center of mobile multimedia services and applications, where the video transmission process must be accomplished and optimized in real-time according to the human's perception, content characteristics, and also context-awareness. Moreover, HCMN environments require instant adaptation of content and resources according to the human's preferences, experiences, or/and interests, due to increasing situational dynamics, such as mobility, bandwidth scarcity, and frequent disconnection.

Recent cross-layer protocols and services have been discussed, designed, and implemented by taking into account QoE and video-awareness. Understanding and modelling user's behaviours, experiences, feelings, and psychological factors when consuming and sharing multimedia content are key issues for HCMN scenarios. For instance, handover decisions must be improved with human-centric polities, real-time video flows can be routed along the path that provide QoE support, or even multimedia application can be adapted to different wireless network conditions or human's preferences. In this way, all HCMN-based proposals achieved better results for multimedia distribution with QoE support, user's satisfaction, and resource optimization, than non-HCMN schemes.

The integration of HCMN into Internet cloud computing environments allows mobile users to have new media experiences that are not possible from their mobile smartphones or tablets. The cloud can be configured to perform a set of important tasks and services for mobile multimedia users and networks, ranging from assessing the video quality level and load balancing to multimedia transcoding and redundancy/error correction schemes. Hence, the user's experience perceived when consuming human-aware videos increase, while the cloud ìvirtuallyì extends the capacity of mobile devices, such as battery, memory, and processing.

Even in Internet mobile cloud services, mobile users still suffer with network congestions, frequent disconnections, and also low video quality experience. For instance, many devices can upload or download the same/similar content (at the same time) from the Internet by using overloaded (and possible high cost) 3G, LTE, or Wi-Fi networks. However, they could cooperate with each other and only one (or few) mobile device could interact with the Internet cloud and request video flows to be locally shared with all neighbours by using Wi-Fi direct, Bluetooth, or even IEEE 802.11p/WAVE in case of vehicular networks. Wireless devices constantly sensing the environment and sharing scarce resources, they could cooperate with each other to maximize the usage of network resources and the human's experiences. These mobile terminals can form a mobile HCMN cloud that offloads the Internet.

Mobile clouds can be dynamically composed of wireless devices located in same area, sharing the same preferences, contexts, and/or videos. Thus, mobile clouds can increase the accuracy of networking protocols or services (enriching with sensing information), while reducing communication delay, spectrum costs, and extending the range of mobile applications. The integration of mobile clouds with Internet clouds must be seamless to bring many benefits for both users and network/content providers. The Internet cloud can enhance mobile devices/clouds by performing high complex or too energy consuming procedures, such as or human-centric video transcoding.

The rest of this section will group HCMN challenges and open issues into cross-layer human, content, and network or device management schemes, covering cross-layer approaches from link to application layers. Regarding human's management approaches, new online, non-intrusive, and QoE video prediction/assessment solutions must be created to measure the quality level of real-time video flows, which have different characteristics as close as possible to human viewers, such as complexity and motion levels, GoP sizes, and CODECs. In this context, standardization bodies and research groups, such as ITU-T SG 12, have been proposing well-structured QoE assessment models. However, in practice, existing solutions have not been discussed and fully implemented in wireless multimedia systems composed of videos with different characteristics, human preferences or experiences, and heterogeneous cooperative mobile devices. Hence, new in-service 2D and 3D video quality assessment solutions must be implemented by taking a set of human perceptual attributes into account, including the overall video quality, environmental characteristics, device features, perceived depth (for 3D videos), and comfort, which, in turn, are the result of technical, social, and psychological factors. Understanding the behaviour or experience of a single user or groups of users is a critical issue for video quality assessment schemes.

Quality of feeling and emotional-based metrics and experiments are expected to be defined in future HCMN proposals, which must be conducted to improve the accuracy of video quality monitoring approaches. Moreover, the use of crowdsourcing (and data collected from social networks) for subjective human studies reduce the time needed, and also increase the performance of new HCMN solutions. In a future multimedia era, users can pay for video services based on a pay as you experience approach, and not a flat rate (or QoS measurements) as currently happens. The human experience can be used to define service-level agreements and other video distribution contracts. Information about the video quality level will be used as input for content creation and network optimization schemes.

Recently, intensive effort has been made by MPEG and other video-related research groups to create new standards and content-aware schemes for multimedia streaming over Internet, such as the MPEG-DASH and MPEG 21. In this way, a problem to be solved is how to code, decode, and transcode video content, by taking the human experience into consideration and making it as adaptable as possible to different network conditions and mobile devices, such as expected in HCMN scenarios. Signal, image, and video processing schemes are time consuming and hard to implement in real-time. Therefore, the description of the video characteristics (in the packet headers or by using addition channels, such as context/name, inter- and inter-frame importance, in/out region-of-interest frame, and motion and complexity levels) is a key issue for allowing real-time videos to be assessed, negotiated, and adapted with high accuracy. The creation (or adaptation) of novel 2D and 3D video codecs adaptable for human experience environments is still an open issue and must be addressed in future systems. Networking techniques can also use the video description to optimize the delivery process.

Internet and mobile cloud computing enriches transcoding techniques with more powerful features, and also human-awareness for HCMN scenarios. However, the video bitrate adaptation alone is not enough to efficiently create new versions of the content for users in wireless heterogeneous networks. Hence, human models, context-awareness, networking statistics, and environmental sensing information must be used to optimize transcoding schemes for HCMNs. The energy consumption of mobile cloud devices can also be improved, by consuming videos transcoded according to their capabilities, such as resolution and battery levels.

The previous approach can also be extended with caching schemes, which is popular in CCN/ICN. One of the benefits of CCN/ICN architectures is the exploitation of in-network caching. However, a video source that streams multiple video formats to a single geographic location (e.g., home, apartment complex building, and VANET) will consume network resources for each different video quality and size. Caches can be placed at edge/mobile clouds and optimized according to different contexts, human models, and network/device resources. A key requirement for improving caching schemes in HCMN scenarios is to integrate human centric experiences, in a way that each user (or QoE video assessment schemes) provides probing and feedback to the video cache service. This information can be used to optimize and selectively cache/transcode videos as needed and deliver the content to mobile devices as human-adapted streams. Such approaches allow the network to generalize multiple geographic regions containing multiple different devices and coalesce to a small number of high quality video streams, which can be cached in-network.

Human and content management information must be used to improve networking approaches with human, application and environment-awareness. HTTP streaming becomes a dominant approach in commercial deployments and must be extended for reporting human experiences and improving the video delivery process. New transport protocols will replace UDP with more advanced HCMN capabilities, especially in multihoming and cooperative environments. Resource reservation, admission control, routing, mobility, queuing, and access control schemes must be extended with human, content and context-aware information. For instance, in wireless networks with high error rates, packet redundancy approaches, such as FEC, must add redundant video packets according to the user perspective, device characteristics, and information about the source of the network congestion process (not only in a black-box manner). Additionally, mobile devices must handoff to new access points according to QoE/human-related factors and not only based on RSSI or network layer metrics. Mobile cloud devices can cooperate with each other to share resources, content, experience, while reducing the system overhead and increasing the user satisfaction and the usage of scarce wireless resources.

The implementation and validation of HCMN solutions are not simple tasks and require many studies and experiments. The interaction between different areas, such as informatics, engineering, visual arts, medicine, and psychology, is crucial to understand the human visual system, behaviour, feeling, and psychological factors. A common sense is that Internet clouds together with mobile clouds will make the implementation of human-centric multimedia mobile systems possible and will enrich them with more advanced, cooperative, and powerful features. Moreover, SDN allows the implementation and validated of HCMN services in wired and wireless networks.

Acknowledgments

This work is supported by the National Council for Scientific and Technological Development (CNPq), as well as the Foundation for Research of the State of Pará (FAPESPA).

We would like to acknowledge the contributions from the following research groups, GERCOM/UFPA (`http://www.gercom.ufpa.br/`), NR2/UFPR (`http://www.nr2.ufpr.br/`), Network Research Lab/USA (`http://nrlweb.cs.ucla.edu/`), and CDS/Switzerland (`http://cds.unibe.ch/`).

References

[Adzic et al. 2011] Adzic, V., Kalva, H., and Furht, B. (2011). A Survey of Multimedia Content Adaptation for Mobile Devices. *Multimedia Tools and Applications*, 51(1):379–396.

[Aguiar et al. 2014] Aguiar, E., Abelem, A., Riker, A., Cerqueira, E., Curado, M., Braun, T., Mu, M., and Zeadally, S. (2014). A Real-time Video Quality Estimator for Emerging Wireless Multimedia Systems. *To appear at Springer Wireless Networks*.

[Aguiar et al. 2012] Aguiar, E., Riker, A., Abelém, A., Cerqueira, E., and Mu, M. (2012). Video Quality Estimator for Wireless Mesh Networks. In *Proceedings of the 20th International Workshop on Quality of Service (IWQoS'12)*, pages 1–9, Coimbra, Portugal. IEEE.

[Andreopoulos et al. 2006] Andreopoulos, Y., Mastronarde, N., and van der Schaar, M. (2006). Cross-layer Optimized Video Streaming over Wireless Multihop Mesh Networks. *IEEE Journal on Selected Areas in Communications*, 24(11):2104–2115.

[Araniti et al. 2013] Araniti, G., Campolo, C., Condoluci, M., Iera, A., and Molinaro, A. (2013). LTE for Vehicular Networking: A Survey. *IEEE Communications Magazine*, 51(5):148–157.

[Bekmezci et al. 2013] Bekmezci, İ., Sahingoz, O. K., and Temel, Ş. (2013). Flying Ad-Hoc Networks (FANETs): A survey . *Ad Hoc Networks*, 11(3):1254–1270.

[Benyamina et al. 2012] Benyamina, D., Hafid, A., and Gendreau, M. (2012). Wireless Mesh Networks Design–a Survey. *IEEE Communications Surveys & Tutorials*, 14(2):299–310.

[Cerqueira et al. 2014] Cerqueira, E., Lee, E., Weng, J.-T., Lim, J.-H., Joy, J., and Gerla, M. (2014). Recent Advances and Challenges in Human-centric Multimedia Mobile Cloud Computing. In *Proceedings of the IEE International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, Hawaii. IEEE.

[Chellouche and Négru 2012] Chellouche, S. A. and Négru, D. (2012). Context-Aware Multimedia Services Provisioning in Future Internet Using Ontology and Rules. In *Proceedings of the 8th International ICST Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous'11)*, pages 466–473, Copenhagen, Denmark. ICST.

[Cisco 2012] Cisco (2012). Cisco Visual Networking Index: Forecast and Methodology, 2011–2016. Technical report, Tech Report.

[Costa et al. 2013] Costa, D., Guedes, L., Vasques, F., and Portugal, P. (2013). Redundancy-Based Semi-Reliable Packet Transmission in Wireless Visual Sensor Networks Exploiting the Sensing Relevancies of Source Nodes. *WSEAS Transactions on Communications*, 12(9):468–478.

[Dey and Abowd 2000] Dey, A. K. and Abowd, G. (2000). Towards a better understanding of context and context-awareness. In *Workshop on The What, Who, Where, When, and How of Context-Awareness*.

[Fajardo et al. 2011] Fajardo, J. O., Taboada, I., and Liberal, F. (2011). QoE-driven and Network-aware Adaptation Capabilities in Mobile Multimedia Applications. *Multimedia Tools and Applications*, pages 1–22.

[Foukalas et al. 2008] Foukalas, F., Gazis, V., and Alonistioti, N. (2008). Cross-layer Design Proposals for Wireless Mobile Networks: a Survey and Taxonomy. *IEEE Communications Surveys & Tutorials*, 10(1):70–85.

[Fu et al. 2013] Fu, B., Kunzmann, G., Wetterwald, M., Corujo, D., and Costa, R. (2013). QoE-aware Traffic Management for Mobile Video Delivery. In *Proceedings of the IIEEE ICC, Workshop on Immersive & Interactive Multimedia Communications over the Future Internet (IIMC)*, Budapest, Hungary. IEEE.

[Gall 1992] Gall, D. J. L. (1992). The MPEG video compression algorithm. *Signal Processing: Image Communication*, 4(2):129–140.

[Gardlo et al. 2014] Gardlo, B., Egger, S., Seufert, M., and Schatz, R. (2014). Crowdsourcing 2.0: Enhancing Execution Speed and Reliability of Web-based QoE Testing. In *Proceedings of the IEEE International Conference on Communications (ICC'14)*, Sydney, Australia. IEEE.

[Georgopoulos et al. 2013] Georgopoulos, P., Elkhatib, Y., Broadbent, M., Mu, M., and Race, N. (2013). Towards Network-wide QoE Fairness Using Openflow-assisted Adaptive Video Streaming. In *Proceedings of the 2013 ACM SIGCOMM Workshop on Future Human-centric Multimedia Networking*, pages 15–20, Hong Kong, China. ACM.

[Ghahfarokhi and Movahhedinia 2013] Ghahfarokhi, B. S. and Movahhedinia, N. (2013). A Personalized QoE-aware Handover Decision Based on Distributed Reinforcement Learning. *Wireless Network*, 19(8):1807–1828.

[Ghareeb and Viho 2010] Ghareeb, M. and Viho, C. (2010). Hybrid QoE Assessment Is Well-Suited for Multiple Description Coding Video Streaming in Overlay Networks. In *Proceedings of the Annual Communication Networks and Services Research Conference (CNSR'10)*, pages 327–333, Montreal, Canada. IEEE.

[Ghareeb and Viho 2011] Ghareeb, M. and Viho, C. (2011). A QoE-based Approach for Multiple Description Coding (MDC) Streaming over Video Distribution Network (VDN). *International Journal of Multimedia Intelligence and Security*, 2(3/4):202–224.

[Greengrass et al. 2009a] Greengrass, J., Evans, J., and Begen, A. (2009a). Not All Packets are Equal, Part I: Streaming Video Coding and SLA Requirements. *IEEE Internet Computing*, 13(1):70–75.

[Greengrass et al. 2009b] Greengrass, J., Evans, J., and Begen, A. C. (2009b). Not All Packets are Equal, Part 2: The Impact of Network Packet Loss on Video Quality. *IEEE Internet Computing*, 13(2):74–82.

[Gu et al. 2005] Gu, T., Pung, H. K., and Zhang, D. Q. (2005). A Service-oriented Middleware for Building Context-aware Services. *Journal of Network and Computer Applications*, 28(1):1–18.

[Gualdi et al. 2008] Gualdi, G., Prati, A., and Cucchiara, R. (2008). Video streaming for mobile video surveillance. *IEEE Transactions on Multimedia*, 10(6):1142–1154.

[Han et al. 2010] Han, L., Park, S., Kang, S.-S., and In, H. P. (2010). An Adaptive FEC Mechanism Using Cross-layer Approach to Enhance Quality of Video Transmission over 802.11 WLANs. *KSII Transactions on Internet & Information Systems*, 4(3):341–357.

[Hanzo et al. 2007] Hanzo, L., Cherriman, P., and Streit, J. (2007). Video Compression and Communications. *IEEE Press*.

[i SCOPE] i SCOPE. i-SCOPE - Interoperable Smart City service Thought an Open Source Platform for Urban Ecosystems. [OnLine] Available: `http://www.iscopeproject.net/`. (Access date: Jun. 2014).

[Immich et al. 2013] Immich, R., Cerqueira, E., and Curado, M. (2013). Adaptive Video-aware FEC-based Mechanism with Unequal Error Protection Scheme. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC'13)*, pages 981–988, Coimbra, Portugal. ACM.

[Imran et al. 2013] Imran, R., Odeh, M., Zorba, N., and Verikoukis, C. (2013). Spatial Opportunistic Transmission for Quality of Experience Satisfaction. *Journal of Visual Communication and Image Representation*, (0):–.

[ITU-R 2002] ITU-R (2002). ITU-R Recommendation BT.500-11. Methodology for the Subjective Assessment of the Quality of Television Pictures. International Telecommunication Union, Geneva, Switzerland, pp. 53–56.

[Jarupan and Ekici 2011] Jarupan, B. and Ekici, E. (2011). A Survey of Cross-layer Design for VANETs. *Ad Hoc Networks*, 9(5):966–983.

[Khan et al. 2009] Khan, A., Sun, L., and Ifeachor, E. (2009). Content Clustering Based Video Quality Prediction Model for MPEG4 Video Streaming over Wireless Networks. In *Proceedings of the IEEE International Conference on Communications (ICC'09)*, pages 1163–1167, Piscataway, USA. IEEE.

[Khan et al. 2010] Khan, A., Sun, L., Jammeh, E., and Ifeachor, E. (2010). Quality of Experience-driven Adaptation Scheme for Video Applications over Wireless Networks. *IET Communications*, 4(11):1337–1347.

[Khorsandroo et al. 2012] Khorsandroo, S., Noor, R. M., and Khorsandroo, S. (2012). A Mobility Framework to Enhance Quality of Experience through Quality of Service. In *Proceedings of the IEEE International Conference on Innovation, Management and Technology Research (ICIMTR2012)*, pages 630–634, Malacca, Malaysia. IEEE.

[Kiani et al. 2013] Kiani, S. L., Anjum, A., Knappmeyer, M., Bessis, N., and Antonopoulos, N. (2013). Federated Broker System for Pervasive Context Provisioning. *Journal of Systems and Software*, 86(4):1107–1123.

[Kritikos et al. 2013] Kritikos, K., Pernici, B., Plebani, P., Cappiello, C., Comuzzi, M., Benrernou, S., Brandic, I., Kertész, A., Parkin, M., and Carro, M. (2013). A Survey on Service Quality Description. *ACM Computing Surveys (CSUR)*, 46(1):1:1–1:58.

[Kulinsk et al. 2013] Kulinsk, D., Burke, J., and Zhang, L. (2013). Video Streaming over Named Data Networking. *IEEE ComSoC MMTC E-Letter*, 8(4):6–10.

[Kumar et al. 2010] Kumar, G. V., Reddyr, Y. V., and Nagendra, M. (2010). Current Research Work on Routing Protocols for MANET: A Literature Survey. *International Journal on Computer Science & Engineering*, 2(3).

[La and Kim 2010] La, H. J. and Kim, S. D. (2010). A Conceptual Framework for Provisioning Context-aware Mobile Cloud Services. In *Proceedings of the IEEE International Conference on Cloud Computing (CLOUD'10)*, pages 466–473, Miami, USA. IEEE.

[Lai et al. 2013] Lai, C.-F., Wang, H., Chao, H.-C., and Nan, G. (2013). A Network and Device Aware QoS Approach for Cloud-Based Mobile Streaming. *IEEE Transactions on Multimedia*, 15(4):747–757.

[Le Gall 1991] Le Gall, D. (1991). MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, 34(4):46–58.

[Library 2014] Library, V. (2014). YUV Video Trace Sequences. [online] http://trace.eas.asu.edu/yuv/ (accessed date: Jun. 2014).

[Lima et al. 2012] Lima, P., Fonseca, T., Rosário, D., Machado, K., and Cerqueira, E. (2012). Monitoramento Ambiental Através de Redes de Sensores Sem Fio e Redes Sociais. In *Proceedings of the Salão de Ferramentas of the 30th Brazilian Symposium on Computer Networks and Distributed Systems (SBRC'12)*, pages 938–945, Ouro Preto, Brazil. SBC.

[Lin et al. 2012] Lin, C.-H., Wang, Y.-C., Shieh, C.-K., and Hwang, W.-S. (2012). An Unequal Error Protection Mechanism for Video Streaming over IEEE 802.11e WLANs. *Computer Networks*, 56(11):2590–2599.

[Lindeberg et al. 2011] Lindeberg, M., Kristiansen, S., Plagemann, T., and Goebel, V. (2011). Challenges and Techniques for Video Streaming over Mobile Ad hoc Networks. *Multimedia Systems*, 17(1):51–82.

[Lu et al. 2011] Lu, Y., Sebe, N., Hytnen, R., and Tian, Q. (2011). Personalization in Multimedia Retrieval: A Survey. *Multimedia Tools and Applications*, 51(1):247–277.

[Luo and Shyu 2011] Luo, H. and Shyu, M.-L. (2011). Quality of Service Provision in Mobile Multimedia - a Survey. *Human-centric computing and information sciences*, 1(1):1–15.

[Ma et al. 2011] Ma, K. J., Bartoš, R., and Bhatia, S. (2011). Review: A Survey of Schemes for Internet-based Video Delivery. *Journal of Network and Computer Applications*, 34(5):1572–1586.

[Macias et al. 2012] Macias, E., Lloret, J., Suarez, A., and Garcia, M. (2012). Architecture and Protocol of a Semantic System Designed for Video Tagging with Sensor Data in Mobile Devices. *Sensors*, 12(2):2062–2087.

[Matos et al. 2012] Matos, R., andCarlos A. F. Marques, N. F. C., Sargento, S., Chakareski, J., and Kassler, A. (2012). Quality of Experience-based Routing in Multiservice Wireless Mesh Networks. In *Proceedings of the IEEE International Conference on Communications (ICC'12)*, pages 7060–7065, Ottawa, Canada. IEEE.

[Mian et al. 2013] Mian, A. U., Hu, Z., and Tian, H. (2013). Estimation of In-service Quality of Experience for Peer-to-Peer Live Video Streaming Systems using a User-centric and Context-aware Approach based on Bayesian Networks. *Transactions on Emerging Telecommunications Technologies*, 24(3):280–287.

[Miao et al. 2012] Miao, Y., Yang, Q., Wu, C., Jiang, M., and Chen, J. (2012). Multicast Virtual Network Mapping for Supporting Multiple Description Coding-based Video Applications. *Computer Networks*, 57(4):990–1002.

[Moorthy et al. 2010] Moorthy, A., Seshadrinathan, K., Soundararajan, R., and Bovik, A. (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.

[Morgenthaler et al. 2012] Morgenthaler, S., Braun, T., Zhao, Z., Staub, T., and Anwander, M. (2012). UAVNet: A mobile wireless mesh network using Unmanned Aerial Vehicles. In *Proceedings of the 3rd International Workshop on Wireless Networking and Control for Unmanned Autonomous Vehicles (Wi-UAV'12)*, pages 1603–1608, Anaheim, USA. IEEE.

[Mu et al. 2009] Mu, M., Cerqueira, E., Boavida, F., and Mauthe, A. (2009). Quality of Experience Management Framework for Real-time Multimedia Applications. *International Journal of Internet Protocol Technology (IJIPT)*, 4(1):54–64.

[Mu et al. 2012] Mu, M., Romaniak, P., Mauthe, A., Leszczuk, M., Janowski, L., and Cerqueira, E. (2012). Framework for the Integrated Video Quality Assessment. *Multimedia Tools and Applications*, 61(3):787–817.

[Naderi et al. 2012] Naderi, M., Rabiee, H., Khansari, M., and Salehi, M. (2012). Error control for Multimedia Communications in Wireless Sensor Networks: A Comparative Performance Analysis. *Ad Hoc Networks*, 10(6):1028–1042.

[Pimentel-Nino et al. 2013] Pimentel-Nino, M. A., Saxena, P., and Castro, M. A. V. (2013). QoE driven adaptive video with Overlapping Network Coding for Best Effort Erasure Satellite Links. In *Proceedings of the 31st AIAA International Communications Satellite Systems Conference (AIAA'13)*, Florence, Italy.

[Pinson and Wolf 2004] Pinson, M. H. and Wolf, S. (2004). A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transactions on Broadcasting*, 50(3):312–322.

[Politis et al. 2012] Politis, I., Dounis, L., and Dagiuklas, T. (2012). H.264/SVC vs. H.264/AVC Video Quality Comparison under QoE-driven Seamless Handoff. *Signal Processing: Image Communication*, 27(8):814–826.

[Puri et al. 2004] Puri, A., Chen, X., and Luthra, A. (2004). Video Coding Using the H. 264/MPEG-4 AVC Compression Standard. *Signal processing: Image communication*, 19(9):793–849.

[Raake and Möller 2011] Raake, A. and Möller, S. (2011). Recent Multimedia QoE Standardization Activities in ITU-Tt SG12. *IEEE COMSOC MMTC E-letter-Special Issue on Quality of Experience issues in Media Delivery*, pages 10–14.

[Rahman et al. 2012] Rahman, M. A., Kim, H.-N., El Saddik, A., and Gueaieb, W. (2012). A Context-aware Multimedia Framework Toward Personal Social Network Services. *Multimedia Tools and Applications*, pages 1–31.

[Rosário et al. 2013a] Rosário, D., Cerqueira, E., Neto, A., Riker, A., Immich, R., and Curado, M. (2013a). A QoE Handover Architecture for Converged Heterogeneous Wireless Networks. *Wireless Networks*, 19(8):2005–2020.

[Rosário et al. 2013b] Rosário, D., Zhao, Z., Braun, T., Cerqueira, E., Santos, A., and Li, Z. (2013b). Assessment of a Robust Opportunistic Routing for Video Transmission in Dynamic Topologies. In *Proceedings of the IFIP Wireless Days conference (WD'13)*, pages 1–6, Valencia, Spain. IEEE.

[Roy et al. 2011] Roy, J., Vaidehi, V., and Sricharan, M. (2011). QoS Guaranteed Integration Methodology for a WLAN–WiMAX Heterogeneous Network. *Computers & Electrical Engineering*, 37(3):261–274.

[Seppanen and Varela 2013] Seppanen, J. and Varela, M. (2013). QoE-driven Network Management for Real-time over-the-top Multimedia Services. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC'13)*, pages 1621–1626, Shanghai. IEEE.

[Seppanen et al. 2013] Seppanen, J., Varela, M., and Sgora, A. (2013). An Autonomous QoE-driven Network Management Framework. *Journal of Visual Communication and Image Representation*, 25(3):565–577.

[Serral-Gracià et al. 2010] Serral-Gracià, R., Cerqueira, E., Curado, M., Yannuzzi, M., Monteiro, E., and Masip-Bruin, X. (2010). An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks. In Osipov, E., Kassler, A., Bohnert, T., and Masip-Bruin, X., editors, *Wired/Wireless Internet Communications*, volume 6074 of *Lecture Notes in Computer Science*, pages 252–263. Springer Berlin Heidelberg.

[Seshadrinathan et al. 2010] Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K. (2010). Study of Subjective and Objective Quality Assessment of Video. *IEEE transactions on Image Processing*, 19(6):1427–1441.

[Setton et al. 2005] Setton, E., Yoo, T., Zhu, X., Goldsmith, A., and Girod, B. (2005). Cross-layer Design of Ad Hoc Networks for Real-time Video Streaming. *IEEE Wireless Communications*, 12(4):59–65.

[Sharabayko et al. 2013] Sharabayko, M. P., Ponomarev, O. G., and Chernyak, R. I. (2013). Intra Compression Efficiency in VP9 and HEVC. *Applied Mathematical Sciences*, 7(137):6803–6824.

[Shehada et al. 2013] Shehada, M., Fu, B., Thakolsri, S., and Kellerer, W. (2013). QoE-based Resource Reservation for Unperceivable Video Quality Fluctuation during Handover in LTE. In *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC'13)*, pages 171–177, Las Vegas, USA. IEEE.

[Sodagar 2011] Sodagar, I. (2011). The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *IEEE Multimedia*, 18(4):62–67.

[Srivastava and Motani 2005] Srivastava, V. and Motani, M. (2005). Cross-Layer Design: a Survey and the Road Ahead. *IEEE Communications Magazine*, 43(12):112–119.

[Taboada et al. 2013] Taboada, I., Liberal, F., Fajardo, J. O., and Ayesta, U. (2013). QoE-aware Optimization of Multimedia Flow Scheduling. *Computer Communications*, 36(15-16):1629–1638.

[Taleb and Ksentini 2012] Taleb, T. and Ksentini, A. (2012). QoS/QoE Predictions-based Admission Control for Femto Communications. In *Proceedings of the IEEE International Conference on Communications (ICC'12)*, pages 5146–5150, Ottawa, Canada. IEEE.

[Tran et al. 2012a] Tran, H. A., Mellouk, A., Hoceini, S., and Augustin, B. (2012a). Global State-dependent QoE Based Routing. In *Proceedings of the IEEE International Conference on Communications (ICC'12)*, pages 131–135, Ottawa, Canada. IEEE.

[Tran et al. 2013] Tran, H. A., Mellouk, A., Hoceini, S., Augustin, B., and Zeadally, S. (2013). A dynamic QoE routing system for user-centric applications. *Transactions on Emerging Telecommunications Technologies*, 24(3):266–279.

[Tran et al. 2012b] Tran, H. A., Mellouk, A., Hoceini, S., and Souihi, S. (2012b). User QoE-based adaptive routing system for future Internet CDN. In *Proceedings of the Computing, Communications and Applications Conference (ComComAp'12)*, pages 90–95, Hong Kong.

[Tsai et al. 2011a] Tsai, M.-F., Huang, T.-C., Ke, C.-H., Shieh, C.-K., and Hwang, W.-S. (2011a). Adaptive Hybrid Error Correction Model for Video Streaming over Wireless Networks. *Multimedia systems*, 17(4):327–340.

[Tsai et al. 2011b] Tsai, M.-F., Shieh, C.-K., Huang, T.-C., and Deng, D.-J. (2011b). Forward-Looking Forward Error Correction Mechanism for Video Streaming over Wireless Networks. *IEEE Systems Journal*, 5(4):460–473.

[van der Schaar et al. 2003] van der Schaar, M., Krishnamachari, S., Choi, S., and Xu, X. (2003). Adaptive Cross-layer Protection Strategies for Robust Scalable Video Transmission over 802.11 WLANs. *IEEE Journal on Selected Areas in Communications*, 21(10):1752–1763.

[Varela and Laulajainen 2011] Varela, M. and Laulajainen, J. (2011). QOE-driven Mobility Management—Integrating the Users' Quality Perception into Network – Level Decision Making. In *Proceedings of the Third International Workshop on Quality of Multimedia Experience (QoMEX'11)*, pages 19–24, Mechelen, Belgium. IEEE.

[Vatolin et al. 2014] Vatolin, D., Moskvin, A., Petrov, O., and Trunichkin, N. (2014). MSU Video Quality Measurement Tool. [OnLine] Available: http://www.download3k.com/Install-MSU-Video-Quality-Measurement-Tool.html. (Access date: Jun. 2014).

[Volkert and Mitschele-Thiel 2012] Volkert, T. and Mitschele-Thiel, A. (2012). Hierarchical Routing Management for Improving Multimedia Transmissions and QoE. In *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'12)*, pages 1–3, San Francisco, USA. IEEE.

[Wang et al. 2013a] Wang, B., Wang, J., and Lu, H. (2013a). Exploiting Content Relevance and Social Relevance for Personalized Ad Recommendation on Internet TV. *ACM Transaction on Multimedia Computing Communication and Applications*, 9(4):26:1–26:23.

[Wang et al. 2013b] Wang, X., Kwon, T. T., Choi, Y., Wang, H., and Liu, J. (2013b). Cloud-Assisted Adaptive Video Streaming and Social-Aware Video Prefetching For Mobile Users. *IEEE Wireless Communications*, 20(3).

[Wang et al. 2004a] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004a). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

[Wang et al. 2004b] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004b). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

[Watkinson 2004] Watkinson, J. (2004). *The MPEG handbook*. Taylor & Francis US.

[Zhang et al. 2013] Zhang, W., Wen, Y., Chen, Z., and Khisti, A. (2013). QoE-Driven Cache Management for HTTP Adaptive Bit Rate Streaming Over Wireless Networks. *IEEE Transaction on Multimedia*, 15(06).

[Zhao et al. 2012] Zhao, Z., Braun, T., Rosário, D., Cerqueira, E., Immich, R., and Curado, M. (2012). QoE-aware FEC Mechanism for Intrusion Detection in Multi-tier Wireless Multimedia Sensor Networks. In *Proceedings of the 1st International Workshop on Wireless Multimedia Sensor Networks (WiMob'12 WS-WMSN)*, pages 697–704, Barcelona, Spain. IEEE.

[Zimmermann 1980] Zimmermann, H. (1980). OSI Reference Model – The ISO Model of Architecture for Open Systems Interconnection. *IEEE Transactions on Communications*, 28(4):425–432.