

QoE in video streaming over wireless networks: perspectives and research challenges

Guan-Ming Su¹ · Xiao Su² · Yan Bai³ · Mea Wang⁴ · Athanasios V. Vasilakos⁵ · Haohong Wang⁶

Published online: 25 August 2015
© Springer Science+Business Media New York 2016

Abstract The deployment of 3G/LTE networks and advancements in smart mobile devices had led to high demand for multimedia streaming over wireless network. The rapid increasing demand for multimedia content poses challenges for all parties in a multimedia streaming system, namely, content providers, wireless network service providers, and smart device makers. Content providers and mobile network service providers are both striving to improve their streaming services while utilizing advancing technologies. Smart device makers endeavor to improve processing power and displays for better viewing experience. Ultimately, the common goal shared by content providers, network service providers, and smart device manufactures is to improve the QoE for users. QoE is both an objective and a subjective metric measuring the streaming quality experience by end users. It may be measured by streaming bitrate, playback smoothness, video quality metrics like Peak to Signal Noise Ratio, and other user satisfaction factors. There have been efforts made to

improve the streaming experiences in all these aspects. In this paper, we conducted a survey on existing literatures on QoE of video streaming to gain a deeper and more complete understanding of QoE quality metrics. The goal is to inspire new research directions in defining better QoE and improving QoE in existing and new streaming services such as adaptive streaming and 3D video streaming.

Keywords Quality of experience · Video streaming · Wireless networks

1 Introduction

Multimedia streaming in all forms (including YouTube) constitutes 58.6 % of peak Internet traffic in North America [1]. In addition, the deployment of 3G/LTE networks and advancements in smart mobile devices (e.g., smartphones and tablets) has led to high demand for multimedia streaming on smart devices. Strategy Analytic forecasts that mobile traffic will increase by 300 % by 2017, and the main driver of this increase is video streaming [2]. FCC's wireless bureau reported that the wireless demand is inevitably going to exceed the available spectrum [3].

Meanwhile, content providers like YouTube and Netflix are continuously increasing their service capacity to provide rich and diverse content to meet user demand. The size of multimedia content is normally larger than other types of content transmitted over the Internet. For example, Netflix recommends customers to have at least 3, 5, and 25 Mbps connection for their standard-definition, high-definition, and ultra-high-definition videos, respectively. According to Google, there are 72 h of video being uploaded to YouTube every minute, and YouTube streams 4 billion hours of video each month. The large volume of

✉ Guan-Ming Su
guanmingsu@ieee.org

¹ Dolby Labs, Sunnyvale, CA, USA

² Computer Engineering, San José State University, San Jose, CA, USA

³ Institute of Technology, University of Washington Tacoma, Tacoma, WA, USA

⁴ Department of Computer Science, University of Calgary, Calgary, AB, Canada

⁵ Department of Computer Science, Electrical and Space Engineering, Lulea University of Technology, 97187 Lulea, Sweden

⁶ TCL Research America, San Jose, CA, USA

video content and high bitrates make video streaming as one of the most resource demanding Internet services.

Both network service providers and content providers are challenged by the growing demand for video streaming. Compare to wired networks, wireless networks are facing additional challenges of reduced bandwidth capacity, interference, and a higher loss rate. The single-path TCP measurements in major U.S. cellular networks in the Boston and Amherst Areas reports that mobile devices are limited to 1.19–18.32 Mbps bandwidth, 39.92–348.02 ms round-trip time (RTT), and a 0.03–0.27 % loss rate [4]. Though the delay and loss rate are tolerable for video streaming, the bandwidth is not sufficient for high quality video streaming experience. Moreover, limited processing power and battery lifetime on smart devices introduce another level of difficulties. A recent measurement study, based on traces of 40,000 clients in a real ISP, revealed that 30 % of YouTube traffic is redundant due to replays [5]. Furthermore, limited memory on mobile devices also causes 10–70 % of redundant traffic during regular playback [6]. Consequently, the perceived streaming quality and Quality of Experience (QoE) are not comparable to that of wired networks.

The increasing demand of unique challenges in mobile video streaming have drawn attention in both industry and academia. In industry, content providers and mobile network service providers are both striving to improve their streaming services while utilizing advancing technologies. For instance, Netflix's utilizes Amazon Web Services (AWS) cloud's computing power and storage space so that its system can dynamically scale according to demand. Network service providers like AT&T are working closely with content providers like Netflix to better deliver the video content to end users. Smart device makers endeavor to improve processing power and displays for better viewing experience. Ultimately, the common goal shared by content providers, network service providers, and smart device manufactures is to improve the QoE for users.

QoE is both an objective and a subjective metric measuring the streaming quality experience by end users. It may be measured by streaming bitrate, playback smoothness, video quality metrics like Peak to Signal Noise Ratio (PSNR), and other user satisfaction factors. There have been efforts made to improve the streaming experiences in all these aspects. For instance, new streaming protocols [7], streaming proxy [8], and cache management [9] have been proposed to improve the streaming bitrate. Cross-layer optimization [10], rate adaptation algorithms [11], buffer management [12], and scheduling algorithms [13, 14] have been suggested to improve the smoothness of playbacks. In terms of video quality perceived by end users, on one hand, researchers and developers are investigating into enhancing common objective quality metrics [e.g., PSNR, Structural

Similarity Metric (SSIM), and Video Quality Metric (VQM)] [15–22]. On the other hand, subjective quality metrics such as Mean Opinion Score (MSO) are being developed to quantify user satisfactory. Moreover, there are also advancements in video codec for better adaptive streaming quality [10].

In order to enhance QoE of video streaming, it is essential to understand all QoE measurements and to determine the key quality metrics as well as their implications in end-user satisfaction. In this paper, we conducted a survey on existing literatures on QoE of video streaming to gain a deeper and more complete understanding of QoE quality metrics. Knowing the QoE of video streaming problem is basically a resource allocation problem, the main contribution of this survey is to group separated QoE metrics into seven categories and analyze their importance and complexity in video source coding and wireless networks. The paper then provides a comprehensive review of efforts made to improve all categories of QoE metrics, based on which we summarize challenges and opportunities in improving QoE of video streaming over mobile networks. The goals are to provide system designers guidelines to select the QoE according to available system resources; and to inspire new research directions in defining better QoE and improving QoE in existing and new streaming services such as adaptive streaming and advanced video representation streaming.

The paper is organized as follows. In Sect. 2, we will provide a comprehensive background review for wireless networks and video codecs, the two key enabling technologies determining the QoE of streaming videos. We will discuss major challenges of transmitting video via streams and providing QoE over wireless networks to mobile devices. In Sect. 3, we will review current and state-of-the-art network solutions and video coding techniques for QoE enhancement. Section 4 summarizes current achievements and challenges in providing QoE of video streaming, followed by an outlook of future research directions. Section 5 concludes the paper.

2 Background

A mobile video streaming system consists of three major components: the video compression component, the wireless network; and the human visual system (HVS). The characteristics of wireless networks and video codec play critical roles in the mobile video streaming services. Different wireless networks provide different bandwidth and transmission reliability; and different video codec provides different compression ratios and robustness to errors. The adaptation of different wireless network and video codec directly impacts the receiving the perception by humans, or

QoE. In this section, we will first look at a overview of wireless networks and video coding to understand how different systems affect the received video quality experience. Then, we will discuss different QoE used in the end-to-end system.

2.1 Wireless and mobile networks

Wireless networks and mobile cellular networks are two types of access technologies, enabling people to access networks without the constraints of cables and wires.

2.1.1 Wireless networks

Wireless networks can be categorized into three types, based on the transmission range of the networks: wireless PAN (Personal Area Network), wireless LAN (Local Area Network), and wireless MAN (Metropolitan Area Network).

The wireless LAN is the most pervasive type of the three, available at home, workspaces, at cafes and airports. Although many standards for wireless LANs were developed in the 1990s, the IEEE 802.11 standard, also known as WiFi, has emerged as the clear market winner. The 802.11 suite has evolved into several standards, including 802.11b, 802.11a, 802.11g, and 802.11n. The 802.11b, an earlier standard, uses the radio frequency band of 2.4 GHz and transmits up to 11 Mbps. Both 802.11a and 802.11g transmit up to 54 Mbps, with 802.11a operating in 5 GHz band and 802.11g using 2.4 GHz band. The more recent 802.11n operates on both 2.4 and 5 GHz bands, uses multiple antennas at the transmitter and receiver, and transmits up to 600 Mbit/s.

The wireless LANs can transmit up to several hundred meters in range. Compared to wireless LANs, wireless PANs have a lower transmission range, with a typical maximum range of 10 meters. A popular wireless PAN standard is IEEE 802.15.1—Bluetooth. The effective range and transmission rate of Bluetooth varies based on radio propagation, antenna and battery conditions. Most Bluetooth applications are indoor, low power communications in short ranges, and transmit up to 1 Mbps.

Compared to wireless PANs and wireless LANs, Wireless MANs can offer a higher transmission range and reach up to 50 km. Wireless MANs provide mobile broadband connectivity within or across metropolitan cities. The IEEE 802.16 network, i.e., WiMAX (Worldwide Interoperability for Microwave Access), is one such standard that allows devices to connect to the Internet through base-stations connected to the main network. WiMAX targets to deliver up to 70 Mbps.

Table 1 compares the transmission ranges and speeds of different wireless access technologies. Note that the

wireless networks can either transmit at longer distances or offer higher rates but not simultaneously. Operating at the upper end of the transmission range results in a lower transmission speed and a higher bit error rate.

2.1.2 Mobile cellular networks

The mobile cellular networks are pervasively used for wide area voice and data communications. The number of cellular subscribers has now surpassed the number of main telephone lines. Since 1980s, the cellular network has gone through several generations, from analog voice in 1G, to digital voice in 2G, to digital voice and data in 3G, and more recently to ultra-broadband speed of gigabit in 4G. Table 2 summarizes the transmission rates and access technologies used in digital cellular networks of 2G, 3G and 4G.

The 1G cellular networks delivered analog voice and the 2G networks delivered digital voice. Digitizing signals allows voice data to be compressed to reduce bandwidth usage and to be encrypted to improve the security and privacy. The dominant 2G system is GSM (the Global System for Mobile Communications), first deployed in the 1990s in Europe. The GSM technology was quickly adopted by countries outside Europe and it became a worldwide success. The transmission speed is comparable to a dial-up link.

Driven by the increasing demand of data traffic from text messaging and video streaming, the 3G networks support digital data service in addition to digital voice. The 3G networks no longer use TDMA, and instead use more advanced technologies, such as CDMA 2000 and UMTS, to support high-quality voice transmission, multimedia applications, text messaging and web browsing. The 3G networks can deliver transmission rates from 500 kbps to 3.1 Mbps in peak performance.

Currently 4G/4G LTE (Long Term Evolution) service is available in every major cellular provider in the US. The 4G LTE networks promise higher bandwidth between 100 and 1000 Mbps, thus can potentially provide higher QoE for streaming applications.

2.1.3 Challenges imposed by wireless and mobile networks

Streaming videos in wireless and mobile networks faces unique challenges caused by the characteristics of the networks.

2.1.3.1 Reduced and variable bandwidth As discussed before, wireless and mobile networks offer lower bandwidths than wired networks. In most scenarios, wireless transmission won't be able to attain the constant speeds listed in Tables 1 and 2, as transmission signals attenuates

Table 1 Comparison of different wireless networks

	Transmission range	Speed (Mbps)
IEEE 802.11b (wireless LANs)	10–100 m	11
IEEE 802.11a/g	10–150 m	54
IEEE 802.11n	20–270 m	600
IEEE 802.15.1 (Bluetooth)	10 m	1
IEEE 802.16 (WiMAX)	50 km	70

Table 2 Comparison of cellular network technologies

	Speed	Technology
2G	Up to 14.4 kbps	TDMA
3G	3.1 Mbps (peak); 500–700 kbps	CDMA2000, EDGE, UMTS
4G	100–1000 Mbps (peak); 3–5 Mbps	WiMAX, LTE

over distance. The limited and variable bandwidth capacity presents a big challenge for bandwidth-demanding video streaming applications.

2.1.3.2 Higher and bursty bit error rates Packet losses could be the result of many factors, including link or node failures, route changes, or bit errors [23], and bit errors could be the result of interference or other characteristics of wireless signals. Wireless networks face the challenge of having to deal with interference caused by other wireless devices or the environment degrading the signal. This occurs on all network frequencies, as transmissions all share the same medium. Licensed bands, such as cellular bands, have restrictions that result in a more controlled environment, compared to unlicensed bands such as WiFi and Bluetooth. The wireless medium is much more susceptible to interference than wired mediums, resulting in a much higher packet loss rate.

2.1.3.3 Larger transmission delays and jitters Real-time interactive video applications are sensitive to transmission delays and variations in packet delays, known as jitters. To ensure continuous and high quality video playback, packets have to arrive at constant intervals. Larger jitters result in jerky and poor quality of videos. The interference, signal fading, and multi-path propagation of wireless transmissions result in larger delays and jitters compared to wired networks. The problem is more serious in mobile networks. When users are roaming from one mobile network to another, the time taken to process mobile handoffs add to both transmission delays and jitters.

2.1.3.4 Limited resources in mobile devices Mobile devices are constrained in computing power, battery life, memory capacity, and display size. These resource limits add more challenges to streaming applications that typically require large bandwidth, fast processing and large buffering.

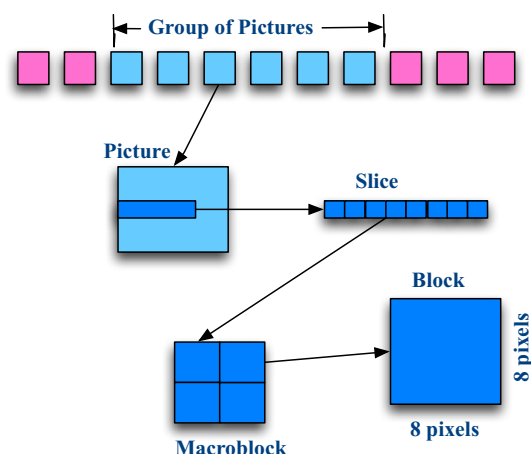
2.2 Video coding

A raw video stream will require a very high data rate to stream, compared to a coded video stream. Coding techniques can be used to compress the raw data, since much of the video data is correlated to adjacent pixels or adjacent frames.

Before we delve into details of spatial and temporal coding, let us examine the structure of a coded video sequence. Figure 1 shows that a video sequence is organized in hierarchies. At the top, the video sequence is divided into groups of pictures (GOP). Each picture is composed of a number of slices—the basic synchronization unit in the sequence. A slice usually consists of a row of 16×16 macroblocks, which has four 8×8 blocks.

2.2.1 Intra-frame coding

Most video coding standards, including H.263, H.264/AVC [24], H.265/HEVC [25], VP8, and VP9, use block-based coding to remove correlations among adjacent pixels

**Fig. 1** Bitstream structure in a video sequence

within the same frame. The intra-frame coding uses an $N \times N$ (e.g., 4×4 , 8×8) block as a unit, and consists of the following steps.

1. Intra-frame prediction: in modern codec, such as H.264 and H.265, block pixels are predicted using the available reconstructed neighboring pixels. The predictions are chosen in the directional fashion for the best fit of local context or minimizing the joint cost of prediction error and prediction overhead. The difference between the original uncompressed pixels and the predicted pixels are processed in the following steps.
2. Discrete Cosine Transform (DCT) to decorrelate the signal so that only a few transformed coefficients are large in magnitude and the rest coefficient values are negligible.
3. Quantization to reduce the magnitude of the transformed coefficients;
4. Applying entropy coding, such as Huffman coding or arithmetic coding, to the quantized coefficients in a zigzag scan order based on the local context. Note that the entropy coding is variable length coding, and synchronization is needed between the encoding and decoding process to locate the correct compressed code words. Thus, there exists a strong decoding dependency to decode consecutive symbols correctly.

Note that in the above process, steps 1, 2, and 4 are invertible and lossless. Step 3 is a lossy non-invertible process, and this is where compression losses occur.

2.2.2 Inter-frame coding

In motion videos, oftentimes the background is kept constant, requiring only the data for moving objects to be transmitted. This can be accomplished by splitting frames into three types: intra frames (I-frames), predictive frames (P-frames), and bidirectional frames (B-frames). An I-frame contains all the data necessary to display the frame, while P and B-frames both depend on previous data to display the frame. P-frames hold the changes between a previous frame and the frame to be displayed, greatly reducing the amount of data to be transmitted. B-frames are extended from I/P-frames in that they hold changes from previous and future frames, which result in even less data being sent. Because of the dependencies from the different types of frame, frames must be encoded with I-frames first, followed by the P-frames that are required by B-frames, and the B-frames.

Figure 2 shows the encoding and decoding order for a sample set of frames. It shows a video sequence of 6 frames, in which frame 1 is coded first, followed by the next P frame, 4, followed by two B frames in between frame 1 and 4, and then followed by the next P frame 6,

and B frame (5) in between frame 4 and 6. The decoding order is the same as the encoding sequence, which is 1, 4, 2, 3, 6, 5. Note that after decoding, the videos are displayed in sequence, 1, 2, 3, 4, and so on.

The difference of display order and decoding order calls for a buffer at the decoder side. For example, frame 2 cannot be displayed until it has decoded frame 4. On the other hand, frame 4 cannot be displayed immediately after being decoded; it will be displayed after both frame 2 and 3 are decoded and displayed.

When encoding P or B frames, the frames that they depend on are called a reference frame; only the differences between P (or B) frame and its reference frame are coded. To account for object movement in the sequence, video coding uses motion estimation to find the best match in the reference frame. Motion estimation process uses macroblock as a unit, searches in a larger window of the reference frame, and identifies the macroblock with the smallest cost (such as differences and/or other encoding overhead) as the best match macroblock. The movement of the macroblock is recorded as motion vectors and will be coded in the bitstreams. Then the differences between the current macroblock and the best match macroblock in the reference frame go through the steps 2–4 as in intra-frame coding. Note that the video content having higher motion often requires a higher bit rate to deliver similar perceived video quality compared to lower motion video. This kind of content dependency provides another dimension of diversity to explore in the video streaming scenario. Similar to the decoding dependency in the intra-frame coding, inter-frame coding also exhibits decoding dependency on the reference frames. A decoding error in the reference frame will propagate within the current frame and to other dependent frames.

2.2.3 Scalable coding

Wireless networks vary in throughput. To keep the QoE of video streaming high, streams must be capable of being scaled by bit rate. SVC (Scalable Video Coding) allows a video to be encoded once and decoded and accessed many times at different rates and ideally at any rate [26]. Possible

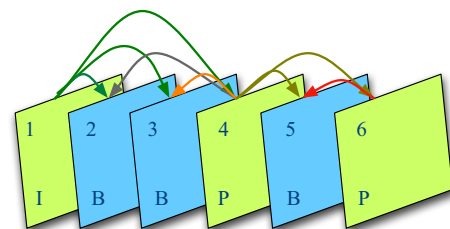


Fig. 2 Encoding order for different frame types

methods of scaling bitrate include scaling the resolution, quality, or frame rate. One example of scalable coding is layered encoding, where a video stream is encoded into multiple layers—a base layer and one or more successive enhancement layers to improve quality [27]. When the network speed varies across devices, layers can be added or dropped to adjust the quality according to the network capability. This allows a video stream to be tuned based on bandwidth and network conditions. Figure 3 shows the encoding and decoding processes of a basic layered encoding scheme.

Non-scalable video coding, as opposed to scalable video coding, is optimized for a single bit rate. When the video content is streamed across multiple connections with different bandwidth, the mileage varies. If the network cannot handle the bit rate of the stream, the decoded videos will have poor quality. However, compared to scalable coding, non-scalable video coding has higher coding gain, i.e., it can typically compress the video better, resulting in smaller bit rate.

2.2.4 Challenges: error resilience

The limited bandwidth and error prone nature of wireless networks presents big challenges to provide high QoE in video streaming. For efficient transmission, a video stream must be encoded to reduce its bit rate by removing redundancies in the video. On the other hand, a video stream depends on its inherent redundancies to recover from lossy transmissions. These two challenges present contradictory requirements in video streaming solutions.

Errors in the coded video stream will accumulate and affect subsequent frames, since some frames are encoded based on previous and future frames. One method to limit frame degradation is by periodically inserting I-frames to

refresh the full video sequence. Incorporating error coding and ARQ (Automatic Repeat Request) techniques can also be used to ensure that the data received is correct, but this will increase bandwidth usage and network delay. Compression efficiency and error handling must be balanced to achieve a good QoE. Figure 4 shows the effect of increased transmission errors on a video frame.

When errors occur in the received data stream, their impact can be reduced by using error concealment techniques. Regions with errors can be corrected by using the pixels surrounding the error region to estimate the actual pixel values. A lost frame can be approximated by replaying the previous frame. By concealing errors, a higher loss rate can be tolerated while maintaining a similar QoE. Figure 5 shows an image with and without concealment applied.

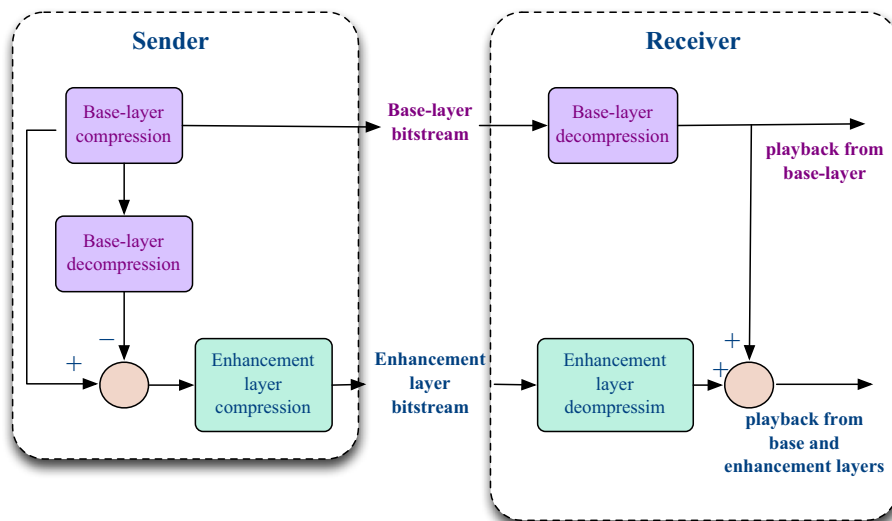
2.3 QoE metric

The wireless network is often designed to optimize the data transmission. As we have seen in the previous section, compressed video exhibits different characteristics from normal data, such as a significantly higher overall bit rate, highly fluctuated bit rate from frame to frame, higher sensitivity to transmission error, and higher sensitivity to delay. Different quality of experience metrics should be designed to address the unique video streaming issues.

We can categorize the mobile video streaming QoE into the following categories according to the computation complexity and applications:

1. *Bit rate* A bit stream with higher bit rate normally provides better perceptual quality than lower bit rate for the same video content from the same video codec. Using bit rate is a convenient way to distinguish the QoE. However, this simple metric is often misleading

Fig. 3 Layered encoding and decoding process



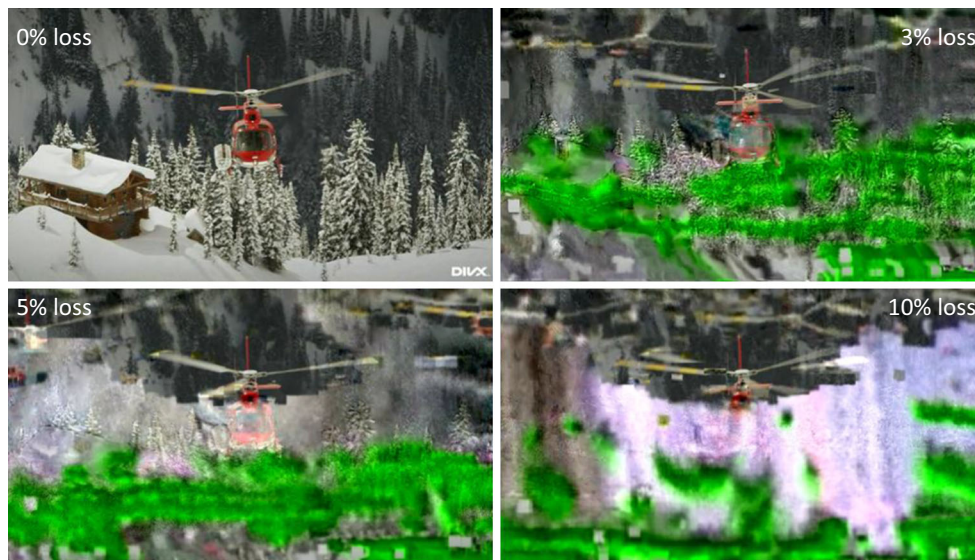


Fig. 4 Example reconstructed video frames after packet losses

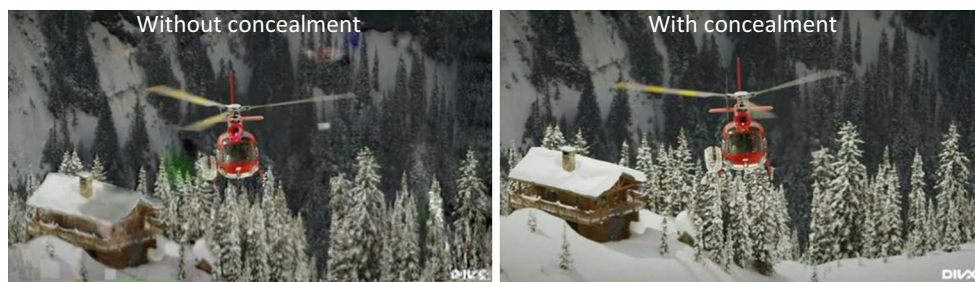


Fig. 5 Example frame shows difference after concealment

as it does not consider the video content complexity, motion, and reflect the dynamics of error-prone wireless network conditions

2. *Playback smoothness* playback smoothness is an important measurement to check how humans perceive the video quality in temporal domain. Playback that is discontinuous and jitters significantly degrades the viewing comfort. It is used to capture the QoE contributed from the wireless network. The dynamics of a wireless network will affect the arrival of video packets and thus affect the timing and correctness to display each corresponding video frame.
3. *Classical objective quality metric* The classical objective quality metric tries to measure the video quality degradation in spatial domain between the decompressed video and the original uncompressed video for each time aligned video pair. The common metric is Peak-to-Signal–Noise Ratio (PSNR), Structural Similarity Metric (SSIM), and Video Quality Metric (VQM). Unlike the bit rate metric, the objective quality metric provides a better QoE metric to address the content complexity and motion. On the other hand,

it takes more computation power to calculate the objective quality metric, and often times it is not feasible to evaluate the quality at the received side owing to the lack of source content.

4. *Subjective quality metric* Subjective quality metrics provides the true QoE since the evaluation is conducted from a group of testing subjects. The metric is the consensus from real human observers. However, conducting subjective testing is both time consuming and economically inefficient. This method cannot be scaled to a large amount of video content with different streaming scenario. A common way to tackle this issue is to develop a mapping function from existing objective quality metrics to the subjective quality metrics with consideration to the features of video content.
5. *Subjective quality metric for scalable video coding* Owing to the highly dynamic network conditions, it is often desired to choose scalable video coding to adapt the bit rates to ensure the playback smoothness. Developing a subjective quality metric for scalable video coding can facilitate the QoE optimization. The

scalable video coding provides several scalabilities, such as spatial, temporal, or PSNR. The subjective quality metric will be a non-linear function to incorporate those scalability parameters.

6. *Subjective QoE metric in lossy environment* The wireless network is not reliable and packet delay and loss often occurs. The video quality degradation should include both the video compression distortion and the communication induced distortion. It is important to model those network impacts to the subjective QoE metric to truly reflect the end-user experience.
7. *Video quality metric for advanced video coding* In recent years, there have been a lot of advanced video coding to provide more realism video experience. One of the examples is 3D video. 3D video has different characteristics compared to the 2D video case since left eye and right eye watch the same scene but from different viewing angles. More importantly, a 3D video should show reasonable depth for the foreground and background. Viewing comfort and fatigue are also important to address. Thus, 3D video needs a different QoE metric for evaluation. Other advanced video coding, such as high dynamic range video, which is in its infant stage, also needs a new video quality metric.

Some QoE metrics are easier to compute, and potentially have the real-time feedback from the received side. They are easy to incorporate into a large scale end-to-end system design. In this case, the entire QoE favors the optimization in the wireless network resource utilization. Some QoE metrics require extensive computation power to calculate to reflect the perceived video quality for video over unreliable channels. Although those QoE metrics provide higher fidelity of measurement, the higher computation complexity also limits the deployment in the practical system. We show the distribution of these QoE metrics in Fig. 6 by analyzing

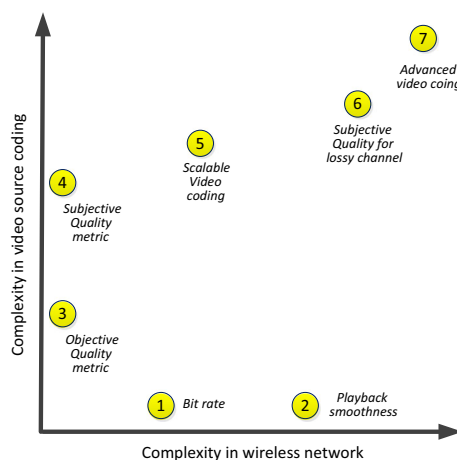


Fig. 6 Computation complexity in different QoE metrics

the required computation complexity in the wireless network domain and in the video compression system domain. Therefore, it is important to understand the characteristics of wireless networks, video codec, and ultimate service goal. Then the system designer can select the best-fit QoE requirement. We summarize the above categorizes, corresponding complexity, and related works in Table 3.

In the following sections, we will discuss how to optimize the selected QoE requirement in different scenario to meet the service goals.

3 Quality of experience model and wireless video streaming

As the traffic volume of video over wireless networks grows, it is becoming increasingly challenging to offer wireless videos with excellent quality of experience (QoE). In recent years, many studies have been focused on analyzing the influencing factors on perceptual quality of videos, development and the application of QoE models to optimize video streaming over wireless networks. Based on the influential factors, we have classified them into seven categories and we will discuss them in each subsection.

3.1 Video bit rate as quality metric

As mentioned in the previous section, a video encoded at higher bit rate normally can provide a better visual quality; though the relationship between the bit rate and the perceived video quality is a content-dependent non-linear function. To design a multi-user video streaming service, using bit rate as a quality measurement often can simplify the system design. Several frameworks have been proposed to optimize the bit rate allocation and bit rate switching.

One approach based on in-network resource management framework, AVIS, is to control the frequency bit rate switching per user via scheduling [7]. AVIS is designed to achieve fairness, stability, and efficiency for multiple DASH (Dynamic Adaptive Streaming over HTTP) video transports over the same base station. It consists of three main components: resource isolation, an allocator, and an enforcer. Resource isolation separates the resource management of adaptive video flows from regular video flows and other data flows. It provides operators with flexibility in resource allocation and management for different kinds of traffic flows, whereas an allocator and enforcer will optimally allocate and schedule bit rates to different adaptive video flows to ensure fairness, stability and high utilization. AVIS deploys utility optimization for resource allocation and enables a balance of optimal bit rate for individual users and average bit rate switches between different users. Therefore, fair resource allocation and good

Table 3 Summary of the QoE category and related works

Category	Complexity in wireless network	Complexity in video coding	Related works
Bit rate	Low	Very low	[7, 9]
Playback smoothness	Medium	Very low	[8, 11–14, 28, 31–34, 36–40]
Classical objective quality metric	Very low	Low	[4, 15–21, 41, 117]
Subjective quality metric	Very low	Medium	[26–33]
Subjective quality metric for scalable video coding	Medium	Medium	[43–50]
Subjective QoE metric in lossy environment	High	High	[22, 42, 59–73]
Video quality metric for advanced video coding	High	High	[74–87]

QoE are obtained at the same time. This is applicable for the transportation of large volumes of adaptive videos over 4G (LTE or i) cellular networks.

Yet in [9], the storage aspect of DASH videos is considered and a QoE-driven cache management is presented. A logarithmic QoE model is built upon experiments relating to the content cache management for HTTP ABR streaming is formulated into a convex optimization (i.e., snapshot problem). Then the Lagrange multiplier method is applied to solve the optimization problem and obtain numerical solutions for the set of playback rates for an individual piece of content cached in the stream engine. Furthermore, three alternative search algorithms (i.e., exhaustive search, Dichotomous-based search, and variable step-size search) can be used to find the optimal number of cached files.

As we can see, using bit rate can simplify the system optimization objective and facilitate the entire end-to-end system resource allocations. On the other hand, this highly simplified QoE metrics often do not address the network influencing factors and the consequent playback issues.

3.2 Playback smoothness as quality metric

The video start-up time often affects the viewers' willingness to continue to watch the video. If the waiting time to get the video playback to start takes too long, the viewers will give up on the viewing. Playback jitters is also an annoying problem for the video streaming. The frequent video pause-and-go often brings unacceptable viewing experience and often makes the viewers to give up the viewing. Therefore, maintaining playback smoothness is also an important quality metric to keep viewers engaged. Packet delay and packet loss in wireless networks might cause a playback jitter and the inherited decoding dependency (e.g., I/P/B frame, decoding order and display order) in the compressed video stream will make the problem even worse. Placing a buffer with an intelligent controller to avoid buffer starvation or overflow can alleviate this problem.

A cross-layer optimization algorithm to improve video quality and video capacity is proposed in [28–30]. The basic

idea is to integrate packet loss visibility models, unequal error protection, and playback buffer regulation. Packet loss visibility modeling is performed at the application layer and estimates the loss of visibility of each video layer in scalable video coding (SVC)-coded videos. Unequal error protection to each video layer is performed at the physical layer to maximize perceptual video quality under packet loss and achieve a better use of link capacity using a joint source-link adaptation. Playback buffer regulation reduces playback buffer starvation and prevents its impacts on video quality such as frame freezes and re-buffering.

A multi-link rate adaptation algorithm for video streaming over multiple wireless access networks, such as WiFi and 3G, is proposed in [11]. The rate adaptation algorithm is implemented in a DASH system in which multiple copies of pre-compressed video with different resolutions and qualities are stored in segments. It requests an appropriate, quality version of video segments based on the current buffer length and available bandwidth. The Markov Decision Process (MDP) was adopted to describe the available bandwidths of different links. In order to optimize the rate adaptation algorithm, a reward function is designed to consider the QoE requirements for video traffic, such as the interruption rate, average playback quality, playback smoothness, and wireless service cost. There are also other studies about rate adaptation. Taking into consideration the starvation probability of playout buffer, continuous playback time, and mean video quality, a scheme [31] adopted an analytical approach to predict the impacts of bit-rate switching on QoE under channel and buffer variations. A scheduling framework to improve fairness, stability, and efficiency of bit-rate switching is proposed in [7]. Both simulations are in a LTE system and prototype implementation in a WiMAX system which demonstrates its effectiveness.

Another approach, called iProxy, is leverage in network caching to lower buffering and start up delay as well as increase bit rates. It also optimizes bit rate switching [8]. iProxy is a mobile video-centric proxy cache that offers better hit rates and streaming quality as compared to state-

of-the-art schemes. It caches video information, rather than video data, and collapses multiple related cache entries into a single one, and thus it improves hit rates while lowering storage costs. Moreover, a dynamic linear rate adaptation scheme ensures high stream quality in face of channel diversity and device heterogeneity. When combined, iProxy yields better bit rates, lower buffering and start up delays. It also optimizes bit rate switching.

A study [12] investigated the impact of buffering interruption on video quality under different network conditions (e.g., low, medium and high bandwidth) and video sources (low and high quality). In particular, users' acceptability with respect to initial loading time, and number of rebufferings and buffering time during video playback is measured, from objective and subjective tests. Their studies have shown that initial loading time of a video stream is less crucial than smoothness of the video playback. It is generally acceptable for a waiting time (sum of initial load time and rebuffering time) of less than 20 s, but intolerable for a waiting time of more than 60 s. Some mathematical models for studying the trade-off between probability of interruption in media playback, and the initial waiting time before starting the playback are constructed in [32]. They concluded that: (1) when arrival rates are slightly larger than the play rate, the minimum initial buffering for a certain level of interruption probability remains bounded as the file size grows, and (2) when arrival rates and play rate match, the minimum initial buffer size scales as the square root of the file size. An evaluation study on HTTP adaptive streaming (HAS) conducted over 3GPP LTE networks is presented in [33]. They defined a set of 3GPP QoE metrics, including HTTP request/response transactions, representation switch events, average throughput, initial playout delay, buffer level, and playlist. A study of rebuffering distribution across users with fixed rate and adaptive streaming was performed. Their research demonstrates that DASH can achieve high video quality with minimal occurrences of rebuffering events, and deliver enhanced QoE to a larger number of LTE clients. A study of QoE for streaming video over LTE networks was conducted in [34]. They defined rebuffering outage capacity to quantify the video service capacity, and proposed QoE-aware adaptive streaming at the radio access network level, which could significantly optimize video capacity.

An improved QoE through application-level scheduling in mobile content distribution network (CDN) was proposed in [13, 35]. The scheduling algorithm leverages the information from the locally available transport and application layer, such as TCP session statistics and content encoding rates, and the information from the mobile infrastructure components, such as congestion levels or subscriber information, to provide a fair distribution of

bandwidth among mobile users. QoE is assessed by the frequency of buffer starvation experienced by each user. In [14], a multilink-based scheduling algorithm to avoid deadline misses is proposed. It divides each video into independent segments. The size of segments varies according to the throughput of available network interfaces (e.g., LAN interface, WLAN interface, 3G interface). Each segment is then transmitted over different networks. Experimental studies with on-demand streaming and live streaming with and without buffering in a controlled network environment and with real world wireless links have shown that the approach efficiently aggregates all available bandwidth, avoids deadline miss and interruption in playback, even when buffer size is limited.

Some other work has been done on the development of QoE evaluation methodologies and performance metrics so as to assist QoE-aware network adaptation and service provisioning. A QoE model that considers the duration of playback interruption and video resolution was built in [36]. A QoE-aware resource management adapts video quality according to user's demand and network resource conditions. An Android-based video server and a video client is designed and implemented for streaming YouTube videos. Different user profiles and resource profiles are tested out. The results show that the QoE-aware resource management can increase QoE satisfaction by 40 percent more than traditional non-QoE service models. In [37], a unicast/multicast system model for multilayer video transportation over wireless networks was developed. In this model, dynamic time slot allocation, transmission rate adaptation, and adaptive pre-drop queue management are integrated for adaptive transmission of different video layers. Delay-bound violation probability is used as the QoE measure.

Lewcio et al. [38] designed and employed a codec changeover in a voice-over-IP system to eliminate session interruption and reduce changeover-related artifacts. Compared to no adaptation, bitrate switching and network handover techniques, the proposed codec changeover scheme produces the best video quality gain in case of packet losses.

A study of managing playout stalls for transmitting multiple video streams from a base station to mobile clients was conducted in [39]. A fast playout lead-aware greedy algorithm is designed for multiplexing videos. The idea is to maximize the minimum 'playout lead' across all videos. The playout lead of a video is the additional duration of time that the video can be played out using only data currently in the client's buffer. Experiments have been performed under different video traces, different pedestrian and vehicular mobility. The results show that the lead-aware greedy algorithm provides a fair distribution of stalls among mobile clients and similar or lower average number

of stalls per client than equal split, and weighted split algorithms.

A reinforcement learning-based HAS client for the delivery of video over mobile networks was designed in [40]. A Q-Learning algorithm is adopted. Two states, available bandwidth and current client buffer filling level are consisted, three QoE factors, average segment quality level, switching behavior of quality levels and video freezes, are used to model environment states and reward functions to help client to select a quality level for the subsequent video segment. Simulation results show that it outperforms by up to 13 % traditional deterministic HAS client.

The playback smoothness QoE metric can reflect more end-user experience owing to the network issues. The playback smoothness problem can be tackled by introducing buffering, which can be solved by many existing rich networking knowledge and techniques. It is also manageable to achieve the goals in network level, such as fairness and efficiency. On the other hand, the playback smoothness QoE metrics only considers the playback issues from the network influencing factors. They do not address the video content complexity/diversity and do not capture the true perceived video quality.

3.3 Classical objective quality metric

As mentioned in previous sections, the relationship between the bit rate and the video quality is a content-dependent non-linear function. To model this non-linear function, one could deploy the common conventional objective quality metric, such as Peak-to-Signal–Noise Ratio (PSNR), Structural Similarity Metric (SSIM), and Video Quality Metric (VQM), to optimize video streaming. These classical objective metrics are measured between the original uncompressed video and the reconstructed video from the compressed bit stream. It is also feasible to introduce the channel distortion into the objective metrics when we consider the end-to-end video transmission.

In particular, PSNR is defined as follows.

$$PSNR = 20 \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right)$$

where MAX_f is the maximum possible pixel value of an image, and MSE is the *Mean Squared Error*. Given a noise-free $m \times n$ monochrome image I and its noisy approximation K , MSE is calculated using the following equation:

$$MSE = \frac{1}{mn} \sum_{s=0}^{m-1} \sum_{t=0}^{n-1} (I(s, t) - K(s, t))^2$$

The SSIM index is a method for measuring the similarity between two images. An initial uncompressed or

distortion-free image is used as reference in the measurement. SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. The SSIM is measured on the block base between a pair of image and the measurement in each block can be done as follows:

$$SSIM(i, k) = \frac{(2\mu_i\mu_k + c_1)(2\sigma_{ik} + c_2)}{(\mu_i^2 + \mu_k^2 + c_1)(\sigma_i^2 + \sigma_k^2 + c_2)}$$

where μ_i is the average value in the block of the original image, μ_k is the average value in the block of the distorted image, σ_i^2 is the variance in the block of the original image, σ_k^2 is the variance in the block of the distorted image, and σ_{ik} is the covariance in the block between the original image and the distorted image, PSNR and MSE are computationally simple, but do not take into account the viewing conditions and the characteristics of human visual perception [41].

The VQM can be used to measure the perceived video quality for various video applications, including direct broadcast satellites (DBS), standard definition television (SDTV), high definition television (HDTV), video teleconferencing (VTC), and wireless or IP-based video streaming systems. It reflects the main impairments including blurring, block distortion, jerky/unnatural motion, noise in luminance and chrominance channels, and error blocks (e.g., transmission errors). A weighted linear combination of all the impairments metrics is used to calculating the VQM rating [41].

An approach to integrate QoE into Media Independent Handover (called QoEHand) in converged heterogeneous wireless networks was proposed in [15]. The QoEHand consists of three components: video estimator, mapping and adaptation. With the cooperation of these three components, it maximizes human experience of video quality and meets the QoE needs of the current applications subject to available resources in IEEE 802.11e/IEEE 802.16 service classes, so that QoE based on SSIM and VQM is largely improved compared to the original Media Independent Handover. In [42], the authors presented a multicast service to improve QoE of video transport over IEEE 802.11 WLAN, where a structured set of collision prevention, feedback and rate adaptation control mechanisms are adopted. QoE is also evaluated by VQM, as in [15].

There are other works that focus on SSIM as a QoE measure. For example, the effect of different network parameters, such as random packet loss, burst packet loss, uniform jitter, and Gaussian jitter on QoE of Skype based on SSIM was investigated in [16]. An interference shaping

scheme in order to reduce QoE variability was proposed in [17]. The main goal is decreasing the peak power of interference bursting transmitters so as to lower packet loss rate and throughput variation. In their work, QoE is assessed by a modified multi-scale structural similarity (HMS-SSIM) index.

A system that combines rate adaption of source and channel coding, CDMA code allocation and power control to improve the QoE of real-time MPEG-4 Fine Granularity Scalability (FGS) video transportation over downlink multicode CDMA networks was developed in [18]. All of these components when coordinated with each other, take into account video quality requirements, code constraints, and power constraints, and achieve an optimal QoE in terms of MSE and PSNR under a resource-limited network. There are also other works on improving classical QoE in terms of PSNR. For instance, in [19], the authors studied the effects of varying channel conditions and resource constraints in WiMAX on users' QoE, including the reserved rate of a video stream at a base station, modulation and coding scheme, distance between base station and mobile station, and tolerable end-to-end delay. PSNR is used as a measure of QoE. Several optimization algorithms to improve fairness and efficiency of video over OFDM (multiuser orthogonal frequency division multiplex) networks were developed in [20]. Video quality improvement is measured using both PSNR and MSE. The authors in [21] pointed out that user-perceived quality of service (QoS) for video delivery over LTE cellular networks had not been well studied. They designed a new OFDMA (orthogonal frequency-division multiple access) scheduling algorithm to address this issue. The algorithm uses weighted round-robin-based radio resource allocation to achieve high system throughput, and a cross-layer optimized system that dynamically adjusts modulation and coding scheme and codec parameters so as to maximize user-perceived video quality, also in terms of PSNR.

3.4 Subjective quality mapped from classical objective metric

Although the objective quality metrics provide a closer approximation of the video quality evaluation compared to the bit rate itself, owing to the highly complex human visual system, the actual perceived video quality is still not well captured using those classical models. Mean Opinion Score (MOS) is a statistical way to measure the perceived video quality from a group of subjects in testing. Researchers have dedicated a great amount of effort to study the correlation between the MOS and the video coding parameters. Note that conducting the subjective testing is very expensive and time consuming; more importantly, it is not practical for massive encoding and

online broadcasting applications. In the video streaming application, it is desired to have a simple method, such as creating a mapping function between the MOS and the easy-to-calculated objective quality metric, to optimize the system performance.

Zhou et al. [43] presented a general media distortion model taking into account not only traditional QoS metrics (i.e., data loss and delay), but also resource availability. Leveraging this model, a utility-based rate allocation scheme (UBRA) was developed for multimedia over a heterogeneous wireless network. Simulation experiments were carried out to validate the proposed model and demonstrate the effectiveness of UBRA. Experiment results show UBRA outperforms drop-tail and AIMD schemes in terms of MOS and PSNR measures.

An instantaneous video quality assessment (IVQA) metric based on video content, such as scene complexity and motion level, was proposed in [44]. The performance evaluation of IVQA is compared to ten existing objective metrics (i.e., PSNR, MSE, SSIM, MSSIM, VSNR, VIF, VIFP, UQI, IFC and MOVIE) and a subjective metrics (MLDS). The running time of IVQA is much stable, which makes this new video quality evaluation model useful for light weight devices such as video camera and mobile devices.

Currently, different metrics are applied in cross-layer approaches to improve QoE [45–49]. A work in [45] was proposed to aim to maximize the utility function at the application layer. The utility function describes the relation between the data transmission rate and perceived video quality in terms of MOS. They proposed an uplink physical resources distribution scheme, which can assure maximum video quality in general and assure better performance for popular videos. In [46], a framework was also proposed to adopt the utility function in order to achieve optimal QoE for all video users under given network conditions. Objective VQM was used for estimating perceptual video quality and mapped to MOS, which is then applied into their SVC mobile video stream adaption at different layers, from link layer to network layer. The approach effectively manages video delivery in both core network and wireless access network. The same authors further improved QoE of SVC mobile video delivery by introducing a priority marking scheme [47]. It mapped SVC layers to different priorities based on data rate and the quality contribution of a particular layer. Rate adaption follows the priorities during the network congestion. In [48], the authors used MOS to represent user satisfaction. MOS score is mapped linearly to video distortion, which is affected by packet loss. They build a distortion model that accurately captures the exact effect of a network packet loss on QoE with a Group of Picture (GOP)-level granularity. Based on this model, an optimal bandwidth allocation minimizes video

distortion under a resource constrained network. In [49], a solution to address the quality controls over multi-access and multi-operator systems, such as IEEE 802.11 and IEEE 802.16, was proposed. It implemented QoE estimation, QoE mapping and QoE adaption in handover periods over multiple layers, including application, session, and network layers. Experimental results on MOS, Structural similarity metric (SSIM), and video quality metric (VQM) show the proposed approach achieves better video quality.

An enhanced adaptive streaming technique for video streaming over LTE networks was designed in [50]. The adaption is based on subjective quality measures and objective estimation. Capacity analysis using client feedback is also integrated into the adaption. Both video characteristics and device information are considered. The proposed technique can support users with better video quality in PSNR (by more than 50–60 %), and reduce the 95-th percentile value of rebuffering percent (by about 4 % with twice as much as load).

3.5 Subjective QoE metric using scalable video coding

Besides exploring the mapping between the subjective metric and the objective metric, what one can do is to directly explore the video content characteristics and construct the subjective metric from those measurements. In the wireless video streaming applications, the system designers can optimize the subjective metric subject to the current channel conditions by dynamically changing the video encoding parameters. The common video encoding parameters are the quantization parameter (QP), the encoding frame rate (frames per second), and the spatial resolution. Encoding a video sequence encoded at higher QP will result in lower bit rate but poor quality owing to the bigger quantization step size. Adjusting the frame rate to a lower value can also reduce the required bit rate as few frames per second are transmitted. Reducing the resolution of a video can also reduce the bit rate as it encodes less number of pixels, though it will cause blurring at the end user side after deploying upsampling to scale back the resolution. Scalable video coding (SVC) can encode one video sequence into one base layer and several enhancement layers. The base layer bit stream is encoded at a low bit rate to accommodate all worst channel conditions. Once there is additional bandwidth available, the encoder side can transmit additional enhancement layer(s) to provide higher PSNR, higher frame rate, and/or higher spatial resolution.

How to quantify the QoE as a function of those parameters in the scalable video coding system is an important issue since it directly affects how system resources can be used efficiently. There have been several

QoE models [51] developed to facilitate the video streaming service with the assumption that the underlying networks provide almost error free transmission condition.

In general, video QoE for SVC codec over error free channel metric consists of three major components: the quality metric from SNR scalability, Q_{CR} , the quality metric from temporal scalability, Q_{CT} , and the spatial resolution scalability Q_{CS} . Q_{CR} is often a function of PSNR or any content dependent features, such as MPEG-4 edge histogram. Q_{CT} is often a function of adopted frame rate and maximal frame rate; and can include other scene dependent parameters, such as MPEG-4 motion activity. Q_{CS} is a function of adopted image dimension and maximal image resolution; and can also include content dependent parameters.

The final model can be constructed by two forms: additive form [52–54] or multiplicative form [55–57]. The generic additive form can be expressed as

$$Q_C = \alpha + Q_{CR} + Q_{CT} + Q_{CS}$$

and the generic multiplicative form can be expressed as

$$Q_C = \alpha + Q_{CR}Q_{CT}Q_{CS}$$

An example of using multiplicative form based on quantization step size, q , and frame rate, f , for SVC is developed in [55]. Let q_{min} be the minimal quantization step size and f_{max} be the maximal frame rate. Then, the maximal required bit rate and corresponding video quality measured in terms of MOS for one given video sequence using q_{min} and f_{max} can be expressed as $R_{max} = R(q_{min}, f_{max})$ and $Q_{max} = Q(q_{min}, f_{max})$, respectively. By given the encoding parameter (q, f), the required bit rate and video quality can be modeled as a function as

$$R(q, f) = R_{max} \left(\frac{q}{q_{min}} \right)^{-a} \left(\frac{f}{f_{max}} \right)^b,$$

$$Q(q, f) = Q_{max} \left(\frac{e^{-c \frac{q}{q_{min}}}}{e^{-c}} \right) \left(\frac{1 - e^{-d \frac{f}{f_{max}}}}{1 - e^{-d}} \right),$$

where R_{max} , Q_{max} , a , b , c , and d are content dependent and often obtained from MOS- (q, f) curve fitting from massive subjective testing, which is time consuming and expensive. We can reduce the parameter, Q_{max} , by taking the normalized QoE model, $\tilde{Q}(q, f) = Q(q, f)/Q_{max}$, since we only care about the relative quality degradation. However, we still have other 5 parameters to estimate. To make solving parameters more practical, it is desired to have a mapping function, $M()$, which converts the features extracted from the video sequence to those model parameters. The features can be residual signal, such as frame difference, displace frame difference, or motion fields related information, such as motion vector magnitude,

motion direction activity, or other measurements [56]. Therefore, the rate and QoE model can be constructed by calculating the video features.

Having the rate and QoE model, we could formulate the video streaming application as an optimization problem to optimize the QoE metric subject to the bit rate constraint.

In the multi-user video streaming scenario, the search space for each (q, f) to obtain the optimal solution grows exponentially. It is desired to simplify the search space to make solution tractable. Based on above rate and QoE model, in [58], a function mapping rate to QoE model is derived as:

$$\tilde{Q}(R) = e^{-\alpha \left(\frac{R}{R_{\max}}\right)^{-\beta} + \alpha}$$

The authors show that parameters α and β are not sensitive to content and can take values from massive regression results. Thus, only R_{\max} is the only content dependent parameter. Note that by given a targeted bit rate, there will be several possible combinations of (q, f) satisfying the rate constraint, R . An algorithm to construct the optimal mapping table between bit rate and (q, f) index is proposed in [58]. In other words, by given a bit rate R , we can determine video quality $\tilde{Q}(R)$ and which enhancement layers to include. The multi-stream video streaming from a central server scenario can be easily formulated as an optimization problem to maximize the quality subject to overall bit rate constraint.

3.6 Adaptive streaming with subjective QoE metric in lossy environment

In most video streaming scenario, especially wireless environment, some video packets will be lost or delayed during the transmission stage owing to the unreliable channels. The video quality suffered from packet loss or delay will degrade the end users' viewing experience. Since the video bitstream has high decoding dependency, any bit error will cause the following bits decoded incorrectly and propagate the error to the rest of streams until an independent decoding unit is met. The end users will observe a lot of undesired artifacts which do not come from compression. Packet delay will cause the frame decoding behind the targeted timeline and cause playback jitter for real-time streaming application. End users will also feel uncomfortable for this kind of uneven playback speed. It is important to develop QoE models to capture both channel condition and video characteristics for video streaming application through lossy channels.

The authors in [59] evaluate several popular VQA models, including as PSNR, SSIM and its variants, VQM, for video streaming over wireless channel with different error patterns and bit error rate. The study shows that there

is a need to develop better VQA model to describe the wireless video streaming QoE metric to match human perception, especially for high performance with low computation complexity.

In general, video QoE over lossy channel metric consists of two major components: the quality metric from video compression parameter as Q_C and the quality metric from transmission module parameter is Q_T . Q_C is often a function of bit rate, frame rate, average motion vector values, quantization parameter. Q_T is often a function of bit rate, frame rate, packet loss rate, density of burst error and burst error duration.

The final model can be constructed in two forms: additive form [60–63] or multiplicative form [64–72]. The generic additive form can be expressed as

$$Q = \alpha + Q_C + Q_T$$

and the generic multiplicative form can be expressed as

$$Q = \alpha + Q_C Q_T$$

There are some constants used in the QoE model and need to be estimated from massive subjective testing.

An example using the multiplicative QoE model for video streaming is discussed in [70, 71]. The basic idea is to represent the model by addressing the contribution from the application layer parameter and the physical layer parameter separately. The video compression related parameters can be the frame rate (FR) and sender bit rate (SBR). The transmission system related can be packet error rate (PER) or other more specific metric in different wireless network. An example used in [70] is shown as follows:

$$Q = \frac{a_1 + a_2 FR + a_3 \ln(SBR)}{1 + a_4 PER + a_5 (PER)^2}$$

Parameter a_1 to a_5 are obtained from curve fitting. Depending on the selected parameter set, the QoE model can be customized for each scenario. To simplify the parameter estimation procedure, a classification based on temporal and spatial features to categorize the video content was conducted. Each category has its own pre-defined parameter. Thus, whenever we need to obtain the QoE model for a new video sequence, we will first extract the features and categorize it. Then, we apply the parameters from the matched category to derive the QoE model.

For the lossless transmission scenario, if the content provider would like to provide a constant QoE video streaming, the sender bit rate can be derived as a function of a required quality level and frame rate [70]. For lossy transmission scenario, the optimal QoE can be adjusted to choose the application parameters according to the channel feedback. In [71], the parameter decision process is done by adopting the fuzzy logic algorithm for video over

UMTS networks. In [42], the authors further apply above QoE model to the cognitive radio networks to allocate network resource by addressing the physical layer parameter as the dropping probability. Similar multiplicative QoE models for Skype video calls consider sender bit rate, frame rate in application layer and packet loss rate, propagation delay and available network bandwidth in physical layer is also derived in [72].

The QoE model can further model the decoder buffer status to reflect the playback jitter. The authors in [73] propose to the following model with addition buffer status metric Q_B :

$$Q = \alpha + Q_C Q_T + Q_T + Q_B$$

In [22], the authors study more complex factors affecting video perceptual quality when a video bit stream is transmitted in the lossy channel, including the length of loss-affected segment (i.e., the error propagation duration after a loss), loss severity (measured by the error in recovering the loss-affected frames), loss location, the number of packet loss, and loss patterns (spread vs. clustered). They found that the joint effect of loss severity, error length and loss number can be captured well by the sum of PSNR drops in all loss-affected frames with some non-linear clipping, where the PSNR drop of a frame is the difference between the PSNR of the encoded frame and that of the reconstructed frame. They also observed that the impact of the loss location can be characterized by a forgiveness factor that decays exponentially with the distance of the last erroneous frame to the end of the clip. Finally the loss pattern effect was captured by a function of the inter-distances between losses, the loss span and the loss number. Based on these findings, they proposed an objective metric for the quality degradation due to packet losses that considers all these factors.

3.7 QoE for advanced video coding

In the past few years, we have witnessed another big revolution of new video viewing experience which is dramatically different from current two-dimensional one. The three dimension (3D) video has extended the horizontal to let the end users enjoy the additional depth information as in the real world. We also expect the new viewing experience will advance from current standard dynamic range video (SDR) with Rec. 709 to high dynamic range video (HDR) to wider color gamut (WCG).

Unlike 2D video which handles a sequence of 2D images along the time domain, 3D video extends to one more dimension to present two views of images, which are captured from two slightly different angles, and are watched by human's two eyes from stereoscopic display. Besides the color and texture information, the perceived

depth and visual comfort are also important factors affecting the stereoscopic viewing experience. Therefore, assessing the quality of QoE measurement for 3D video is complex and non-intuitive. Deploying 3D video streaming has more challenges than the conventional 2D video. From the source coding's point of view, the 3D video codec explores the inter-view and synthesis prediction to reduce the required bit rate. In other words, 3D video bitstream has more decoding dependency. From channel's perspective, transmitting 3D bitstream still needs larger bandwidth compared to traditional 2D video. More specific and advanced error protection methods to protect bitstream and more efficient approaches to utilize network resources are needed to meet QoE requirements.

Two-view stereo video coding is considered the simplest 3D video coding method that each view is encoded independently. This approach can deploy the 3D stream over networks to end users quickly without changing entire video streaming infrastructure. The drawback is the need of double required bandwidth in the network to accommodate left view and right view bitstreams. Asymmetric stereo video coding is proposed to alleviate the bit rate consumption. The asymmetric stereo video coding adopts different coding parameters in each view, such as different spatial resolution, different temporal frame rate, and different signal-to-noise (SNR) [74–78]. The spatial asymmetry in stereo coding [75, 76] is based on the binocular suppression theory that a good quality stream for one view with a smaller spatial resolution in another view (which requires less bit rate) can provide equivalent QoE compared to same resolution. Temporal asymmetry [78] is to provide an unequal frame rate for both views: one view has a normal frame rate and the other view has lower frame rate. The subjective testing shows temporal asymmetry is effective for slow motion video sequences. SNR asymmetry is the easiest method to deploy since no extra spatial upsampling or temporal frame rate conversion is needed in the decoder side. Study in [74] shows that a threshold exists to reduce the required SNR in the non-dominant eye to achieve same QoE with same level of QoE in symmetric coding.

Multi-view video provides a new visual experience that the end users can watch a scene from different viewpoints in order to achieve more interactive experience. To achieve this goal, the bitstreams should contain sufficient views to the end users and end users will extract the needed stereo pair from the combo bitstreams for watching. H.264 multiple view coding (MVC) [79] extends two-view stereo coding into a more generic scheme to support multiple views. For the MVC streaming application, one important issue is how to allocate bits to each view so the end users can have the best viewing experience. One could set up the objective to minimize the overall distortion summed up

over all possible views [80] or to minimize the maximal distortion in each individual view.

Although MVC can provide around 20 % bit rate saving compared to individual view coding, there is a great desire to further reduce the bit rate especially when the required number of views grow. Multi-view video plus depth (MVD) codec is one of the solutions to achieve this goal. The basic principle of MVD is to transmit few views with color/texture along with depth information in the bitstream. For the views not included in the bitstreams, the decoder will use the depth-image-based-rendering (DIBR) [81, 82] approach to generate the synthesis views. In the MVD streaming scenario, the system designer should consider the QoE for both types of view: the encoded views and the synthesized views. Note that the QoE of synthesized views highly depends on the QoE of reconstructed color/texture and depth information encoded views. In [83–85], QoE of the synthesized view can be modeled as a function of the QoE parameters used in the color/texture and depth pictures from both left and right views. Having the QoE model for the synthesized view, we could formulate the MVD streaming system as an optimization problem to optimize the required QoE.

Owing to the unique features of 3D video, we can further deploy advanced methods to improve the QoE video streaming. One of the methods is the multiple description coding (MDC). MDC is a source compression technique to represent the video signal into multiple descriptions. Each description can be decoded independently and provide the baseline video quality. Whenever the end users receive more descriptions, the QoE can be improved further. Naturally, 3D video fits into this architecture since left view and right view can be seen as an independent description. Several schemes based on MDC to optimize distortion are proposed in [86, 87].

4 Open research issues

The QoE driven video mobile streaming system consists of three major components: (1) the video source compression; (2) the wireless network; and (3) the human visual system (HVS). The required bandwidth to provide satisfactory video quality is very high compared to other types of network traffic. To reduce the amount of video traffic, highly efficient video compression is needed to remove the redundancy in the video signal. The redundancy removal process deployed in the video compression system introduces higher traffic fluctuation and higher decoding dependency owing to independent and dependent contexts. Thus, a compressed bit stream complicates the rate allocation for the channel and requires higher error protection for transmission. As the demand of video streaming

applications over wireless networks grows rapidly, the required bandwidth to support all of the requested users quickly approaches/exceeds the network capacity. Intelligent resource allocation and scheduling are required to fully utilize the network resources and maintain the link reliability to achieve the service goals. In the entire video ecosystem, HVS plays the most important role since the final video bit streams are perceived by human eyes and recognized by human brains. The quality metric should be designed to measure how HVS judges the viewing experience.

From reviewing state-of-the-art network solutions and video coding techniques for QoE enhancement, we identified four major challenges for a QoE driven mobile streaming video. The challenges and their associated opportunities are summarized in the rest of this section.

4.1 Efficient quality metrics

It is essential to construct efficient quality metrics that truly reflect the viewing experience and facilitate the system resources adaptation in both video coding and wireless network.

HVS is a highly non-linear and high dimension system. There have been a lot of efforts towards understanding the video viewing experience in the baseband signal format before compression, namely, the required color space for better color representation and codeword allocation, including bit depth, for dynamic range representation within a scene. With the introduction of video compression, there are more factors affecting the final viewing experience, especially how HVS ranks the artifacts induced by the lossy compression. The researchers are still endeavor to understand how the compression affects the QoE in more details. The task becomes much more difficult when the video signal is impaired during the lossy transmission over wireless networks, since the dimension of this problem expands exponentially to include both source and channel side. The random nature of the wireless networks and the corresponding error propagation exhibited in the video decoding process makes the construction of QoE even more difficult.

As we have seen in Sect. 3, when the existing QoE models approximate the perceived video quality better, the required computation load becomes more expensive. And many models need offline training with iterative encoding/decoding and evaluation, which makes the system not practical for real-time streaming. In addition to the computation complexity, when a QoE model gets complex in the video source coding side, the existing QoE model starts to lack the ability to consider the highly dynamic network conditions, thus the ability to adjust network resources and parameters becomes weaker. In general, a desired QoE

model should have low computation complexity, be scalable for large-scale system, and provide clear tuning points on how to adjust the parameters in the video coding systems and the wireless networks resources.

4.2 Streaming system optimization

Often a new streaming system is proposed to offer either a new type of streaming service or improving certain quality metrics [29, 30, 35, 88–93]. Streaming system researchers and developers shall consider all seven categories of QoE metrics (defined in Sect. 2.3) and determine the target QoE before designing the system. However, this is a challenging task since there may be conflicts posted by QoE metrics. For example, improving PSNR may lead to higher bitrate. Therefore, the best choice of the target QoE deserves attention and requires careful engineering to achieve the optimal viewing experience.

More specifically, a QoE model may consist of multiple parameters contributing to the final quality measurement. Note that those QoE parameters may not be straightforward and mapped to parameters used in the streaming system. To facilitate the performance tuning in the system design, it is important to find the mapping between these parameters in QoE model and the corresponding parameters in the streaming system. To optimize the QoE perceived by each end user, the QoE-driven video streaming system should determine how to adjust the parameters in both video coding system and mobile networks. Some parameters can be adaptive in real time; some can only be done in a longer period; and some can be only adjusted for each connection section. How to dynamically adjust parameters in different time scale to achieve desired QoE is a big challenge. Since the network resources are limited and are shared by multiple users, a more design challenge is to equip an efficient and fair resource management system to provide satisfactory QoE to each user.

4.3 QoE in error-prone channel

Compared to wired networks, wireless networks constitute many error-prone channels. In the past decades, there have been several techniques developed to address the unique characteristics of video over error prone channel. For example, the unequal error protection scheme (either via Automatic Repeat-reQuest (ARQ) or Forward Error Coding (FEC)) is deployed to protect more important syntaxes of video components with stronger protection to achieve a statistical gain.

Unequal error protection (UEP) is a special form of forward error correction (FEC). It protects more important data with stronger forward error correction code. The use of UEP in protecting data transmission over noisy channel

was initially proposed in [94]. The applications of UEP in video streaming exploit either the coding technique or the importance of video packets.

Coding Techniques: In general, any coding technique that generates code at an arbitrary code rate can be used to provide unequal protection. Common coding techniques for UEP include rate-compatible convolutional codes (RCPC) [95], low-density parity-check (LDPC) codes [96], growth codes [97], expanding window fountain codes (EWFC) [97], and Raptor codes [98]. While most of the existing proposals are applications of UEP in the application layer, the UEP protection scheme can also be incorporated in the physical or transport layer at the bit level [99]. The performance of different coding techniques have been analyzed and presented in [100].

Importance of video packets: A video encoder encodes a video sequence into a series of video packets for transmission. UEP provides different levels of protections to video packets according to a specific importance measure. A basic importance measure of a video packet is the position of the packet in the layer dependency structure, i.e., a reference picture is more important than a dependent picture [101]. Data partitioning posed by video encoders (e.g., separation of header data, intra-picture predicted macroblocks), and inter-picture predicted macroblocks) may be used as an importance measure [102, 103]. The bitrate of the encoded video frames or slices can also be used as an importance measure [97, 98, 101]. For example, the larger video frame contains more information and is considered more important. Other importance measures have been explored are playback deadlines of slices [104], the length of the dependency chain among frames during encoding [105], and motion energy (defined as macroblock size times the motion vector size) [106].

In general, existing approaches explore the possibility of error concealment by exploring the high redundancy in the spatial and temporal domain are proposed. If the mobile video streaming system is equipped with those mechanisms, the QoE model can be refined to address those error resilience and error concealment tools. Consequently, the corresponding streaming strategy can be also refined to meet the QoE model.

4.4 QoE model for advanced video systems

Advanced video systems have become very hot topics in recent years. Many of them have and will come to the market very soon. However, their QoE models are not well defined and the corresponding QoE driven solutions remain future research work.

- 3D video system is one of the advanced video systems [107, 108]. The newly introduced depth information

expands the dimension of QoE model. In addition, different stereoscope capture system, display system (such as parallax barrier and lenticular display), compression presentation formats (such as multi-view system and multi-view plus depth system), have different unique QoE issues to tackle..

- Screen content coding [109] is the new coding tool defined in H.265/HEVC. It is commonly used in the screen sharing applications, such as remote team collaboration environment. A new QoE model is needed to define how the end users perceive the visual quality.
- Ultra High definition TV (UHDTV) [110] brings the much bigger viewing areas (4 times larger than the conventional HDTV) and wider color space (2 times of color in Rec. 2020 than conventional Rec. 709) [111]. A QoE model to measure and quantify the viewing experience/impact under different spatial resolution (especially higher spatial resolution) and different colorimetry (especially the much wider color gamut) is needed (e.g., ΔE 2000 [112]).
- High frame rate video: the current video frame rate (24–30 fps) often introduces flicker and motion blur, especially for high motion scenes, such as sport broadcasting and action movies. The motion blur can be alleviated by introducing the high frame rate video system [113]. It is worthy to notice higher frame rate needs higher bandwidth and higher system operating frequency. An effective QoE model to capture the viewing experience benefited from high frame rate service is important as well.
- High dynamic range video (HDR) [114] brings another brand new viewing experience by moving the display referred video presentation to scene referred video presentation. In the HDR ecosystem, the dynamic range of the video signal is significantly expanded to range almost near the nature scenes (thousands of nits, or cd/m^2). The viewing experience is dramatically different from the standard dynamic range (SDR) system, which only provides hundreds of nits. Specific HDR video codecs and HDR image processing pipeline are currently under development to handle the HDR video by inventing new coding/processing tools to address the unique issues in HDR content. For example, instead of existing BT. 1886 signal, a new signal format, namely, perceptual quantizer [115], is introduced to encode the HDR baseband signal for better higher dynamic range signal handling. New types of coding/processing artifact through the streaming service are expected. It is important to develop a new QoE model (such as HDR-VDP [116]) to address this new viewing experience and thus provides a good guideline when and how to provide HDR streaming service.

5 Conclusions

In this paper, we categorize the QoE driven mobile video streaming into several different types according to the computation complexity and fidelity to the final perceived video quality. We also discuss different system designs that optimize different QoE metrics. We observe that there is a trade-off between the accuracy of QoE metrics and system resources utilization. A QoE model which requires less computation complexity often indicates lower accuracy of true perceptual quality, but leads to easier system design to support a large scale of video streaming services. The simpler QoE model facilitates the wireless network resources utilization in a more efficient way. A complex QoE model can approximate the true perceptual video quality better, but the required computation power and process to obtain QoE value couldn't scale a video streaming system large enough. The complex QoE model facilitates the parameters adaptation in the video source coding part.

We also discussed the challenges and opportunities for the QoE driven mobile video streaming. The research trend is to continue to improve the QoE accuracy while lowering the computation complexity of calculating the QoE, and design a system to optimize the defined QoE metric. The ultimate goal is to have a QoE driven video streaming system which can fully utilize the network and video coding resources efficiently and provide the best visual quality. We also briefly review the state-of-the-art video system and the potential to set up the QoE driven streaming system.

References

1. Sandvine Intelligent Broadband Networks. (2012). *Global internet phenomena report: 2H 2012*. http://www.sandvine.com/downloads/documents/Phenomena_2H_2012/Sandvine_Global_Internet_Phenomena_Report_2H_2012.pdf
2. Strategy Analytics. (2013). Handset data traffic to grow over 300 % by 2017 to 21 Exabytes. <https://www4.strategyanalytics.com/default.aspx?mod=pressreleaseviewer&a0=5385>
3. CNN. *Sorry, America: Your wireless airwaves are full*. http://money.cnn.com/2012/02/21/technology/spectrum_crunch/
4. Chen, Y.-C., Lim, Y.-S., Gibbens, R. J., Nahum, E. M., Khalili, R., & Towsley, D. (2013). A measurement-based study of multiPath TCP performance over wireless networks. In *Proceedings of ACM internet measurement conference (IMC)* (pp. 455–468), October 23–25. Barcelona, Spain.
5. Zohar, E., Cidon, I., & Mokryn, O. (2011). The power of prediction: Cloud bandwidth and cost reduction. In *Proceedings of ACM SIGCOMM conference* (pp. 86–97), August 15–19. Toronto, Canada.
6. Liu, Y., Li, F., Guo, L., Shen, B., & Chen, S. (2013). Effectively minimizing redundant Internet streaming traffic to iOS devices. In *Proceedings of 32nd IEEE conference on computer communications (INFOCOM)* (pp. 250–254), April 14–19. Turin, Italy.

7. Chen, J., Mahindra, R., Khojastepour, M. A., Rangarajan, S., & Chiang, M. A scheduling framework for adaptive video delivery over cellular networks. In *Proceedings the 19th annual international conference on Mobile computing and networking* (pp. 389–400). Miami, FL, USA.
8. Shen, S.-H., & AKella, A. An information-aware QoE-centric mobile video cache. In *Proceedings of the 19th annual international conference on mobile computing & networking* (pp. 401–412). Miami, FL, USA.
9. Zhang, W., Wen, Y., Chen, Z., & Khisti, A. (2013). QoE-driven cache management for HTTP adaptive bit rate streaming over wireless networks. *IEEE Transactions on Multimedia*, 15(6), 1431–1445.
10. The Nielsen Company. (2013, September) “Binging” is the new viewing for over-the-top streamers. <http://www.nielsen.com/us/en/newswire/2013/binging-is-the-new-viewing-for-over-the-top-streamers.html>
11. Xing, M., Xiang, S., & Cai, L. (2012). Rate adaptation strategy for video streaming over multiple wireless access networks. In *Proceedings IEEE global communications conference* (pp. 5745–5750).
12. Pessemier, T. D., Moor, K. D., Joseph, W., Marez, L. D., & Martens, L. (2013). Quantifying the influence of rebuffering interruptions on the user’s quality of experience during mobile video watching. *IEEE Transactions on Broadcasting*, 59(1), 47–61.
13. Yousaf, F. Z., Liebsch, M., Maeder, A., & Schmid, S. (2013). Mobile CDN enhancements for QoE-improved content delivery in mobile operator networks. *IEEE Networks*, 27(2), 14–21.
14. Evensen, K., Kasparl, D., Griwodz, C., Halvorsen, P., Hansen, A. F., & Engelstad, P. (2011). Improving the performance of quality-adaptive video streaming over multiple heterogeneous access networks. In *Proceedings the second annual ACM conference on multimedia systems* (pp. 57–68). Santa Clara, CA.
15. Rosário, D., Cerqueira, E., Neto, A., Riker, A., Immich, R., & Curado, M. (2013). A QoE handover architecture for converged heterogeneous wireless networks. *Wireless Networks*, 19(8), 2005–2020.
16. Liotta, A., Druda, L., Menkovski, V., & Exarchakos, G. (2012). Quality of experience management for video streams: The case of Skype. In *Proceedings of the 10th international conference on advances in mobile computing and multimedia* (pp. 84–92). Bali, Indonesia.
17. Singh, S., Andrews, J. G., & De Veciana, G. (2012). Interference shaping for improved quality of experience for real-time video streaming. *IEEE Journal on Selected Areas in Communications*, 30(7), 1259–1269.
18. Han, Z., Su, G. M., Kwasinski, A., Wu, M., & Ray Liu, K. J. (2006). Multiuser distortion management of layered video over resource limited downlink multicode-CDMA. *IEEE Transactions on Wireless Communications*, 5(11), 3056–3067.
19. Vishwanath, A., Duttap, P., Chetlurb, M., Guptab, P., Kalyanaraman, S., & Ghosh, A. (2009). Perspectives on quality of experience for video streaming over WiMAX. *ACM Mobile Computing and Communications, Review*, 13(4), 15–25.
20. Han, G.-M. Z., Wu, M., & Liu, K. J. (2006). A scalable multiuser framework for video over OFDM networks: Fairness and efficiency. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(10), 1217–1231.
21. Luo, H., Ci, S., Wu, D., Wu, J., & Tang, H. (2010). Quality-driven cross-layer optimized video delivery over LTE. *IEEE Communication Magazine*, 48, 102–109.
22. Liu, T., Wang, Y., Boyce, J. M., Yang, H., & Wu, Z. (2009). A novel video quality metric for low bit-rate video considering both coding and packet-loss artifacts. *IEEE Journal of Selected Topics in Signal Processing*, 3(2), 280–293.
23. Liao, Y., & Gibson, J. D. (2009). Rate-distortion based mode selection for video coding over wireless networks with burst losses. In *IEEE Proceedings 17th Int’l packet video workshop* (pp. 1–10).
24. Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), 560–576.
25. Sullivan, G. J., Ohm, J., Han, W.-J., & Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1649–1668.
26. Wang, Y., & Zhu, Q. (1998). Error control and concealment for video communication: A review. *Proceedings of the IEEE*, 86, 974–997.
27. Saporillaand, D., & Ross, K. W. (2000). Optimal streaming of layered video. In *Proceedings IEEE INFOCOM 2000* (pp. 737–746).
28. Khalek, A. K., Caramanis, C., & Heath, R. W., Jr. (2012). A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection. *IEEE Journal on Selected Areas in Communications*, 30(7), 1157–1171.
29. Shen, Z., Luo, J., Zimmermann, R., & Vasilakos, A. V. (2011). Peer-to-peer media streaming: Insights and new developments. *Proceedings of the IEEE*, 99(12), 2089–2109.
30. Zhou, L., Zhang, Y., Song, K., Jing, W., & Vasilakos, A. V. (2011). Distributed media services in P2P-based vehicular networks. *IEEE Transactions on Vehicular Technology*, 60(2), 692–703.
31. Xu, Y., Zhou, Y., & Chiu, D. (2014). Analytical QoE models for bit-rate switching in dynamic adaptive streaming systems. *IEEE Transactions on Mobile Computing*, 1536–1233.
32. ParandehGheibi, A., Médard, M., Ozdaglar, A., & Shakkottai, S. (2011). Avoiding interruptions—A QoE reliability function for streaming media applications. *IEEE Journal on Selected Areas in Communications*, 29(5), 1064–1074.
33. Oyman, O., & Singh, S. (2012). Quality of experience for HTTP adaptive streaming services. *IEEE Communication Magazine*, 50(4), 20–27.
34. Singh, S., Oyman, O., Papathanssiou, A., Chatterjee, D., & Andrews, J. (2012). Video capacity and QoE enhancements over LTE. In *Proceedings 2012 IEEE international conference on communications* (pp. 7071–7076). Ottawa, Canada.
35. Jin, Y., Wen, Y., Shi, G., Wang, G., & Vasilakos, A. V. (2012). CoDaaS: An experimental cloud-centric content delivery platform for user-generated contents. In *IEEE international conference on computing, networking and communications (ICNC)* (pp. 934–938).
36. Yavari, A., Lungaro, P., & Segall, Z. (2013). Network efficient resource management for mobile streaming-based on quality of experience. In *2013 international conference on ICT convergence*, Jeju Island, Korea.
37. Du, Q., & Zhang, X. (2010). Statistical QoS provisionings for wireless unicast/multicast of multi-layer video streams. *IEEE Journal on Selected Areas in Communications*, 28(3), 420–433.
38. Lewcio, B., Möller, S., & Vidale, P. (2011). A method for seamless Codec changeover during active video calls. In *Proceedings IEEE 16th international workshop on computer aided modeling and design of communication links and networks* (pp. 133–137). Kyoto, Japan.
39. Dutta, P., Seetharam, A., Arya, V., Chetlur, M., Kalyanaraman, S., & Kurose, J. (2012). On managing quality of experience of multiple video streams in wireless networks. In *Proceedings IEEE INFOCOM 2012* (pp. 1242–1250). Orlando, FL, USA.

40. Claeys, M., Latré, S., Famaey, J., & Turck, F. D. (2014). Design and evaluation of a self-learning HTTP adaptive video streaming client. *IEEE Communications Letters*, 18(4), 716–719.
41. Chikkerur, S., Sundaram, V., Reisslein, M., & Karam, L. J. (2011). Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting*, 57(2), 165–182.
42. Jiang, T., Wang, H., & Vasilakos, A. V. (2012). QoE-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 30(7), 1215–1224.
43. Zhou, L., Wang, H., Lian, S., Zhang, Y., Vasilakos, A. V., & Jing, W. (2011). Availability-aware multimedia scheduling in heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, 60(3), 1161–1170.
44. Liotta, A., Mocanu, D.C., Menkovski, V., Cagnetta, L., & Exarchakos, G. (2013). Instantaneous video quality assessment for lightweight devices. In *International conference on advances in mobile computing & multimedia*, Vienna, Austria, December 2013.
45. Essaili, A. E., Steinbach, E., Munaretto, D., Thakolsri, S., & Kellerer, W. (2011). QoE-driven resource optimization for user generated video content in next generation mobile networks. In *Proceedings 18th IEEE international conference on image processing* (pp. 913–916). Brussels, Belgium.
46. Fu, B., Kunzmann, G., Wetterwald, M., Corujo, D., & Costa, R. (2013). QoE-aware traffic management for mobile video delivery. In *Proceedings 2013 IEEE international conference on communications workshops* (pp. 652–656). Budapest.
47. Fu, B., Staehle, D., Kunzmann, G., Steinbach, E., & Kellerer, W. (2013). QoE-aware priority marking and traffic management for H.264/SVC-based mobile video delivery. In *Proceedings the 8th ACM workshop on performance monitoring and measurement of heterogeneous wireless and wired networks* (pp. 173–180). Barcelona, Spain.
48. Goudarzi, P., & Hosseinpour, M. (2010). Video transmission over MANETs with enhanced quality of experience. *IEEE Transactions on Consumer Electronics*, 56(4), 2217–2225.
49. Jailton, J., Carvalho, T., Valente, W., Natalino, C., Francês, R., & Dias, K. (2013). A quality of experience handover architecture for heterogeneous mobile wireless multimedia networks. *IEEE Communications Magazine*, 51(6), 152–159.
50. Foerster, J., Oyman, O., Liao, Y., Rehan, M., & Taie, W. (2012). Enhanced adaptive streaming over LTE-advanced wireless networks. In *Proceedings 46th Asilomar conference on signals, systems and computers* (pp. 915–919). Pacific Grove, CA.
51. Lee, J.-S., De Simone, F., Ebrahimi, T., Ramzan, N., & Izquierdo, E. (2012). Quality assessment of multidimensional video scalability. *IEEE Communications Magazine*, 50, 38–46.
52. Feghali, R., Speranza, F., Wang, D., & Vincent, A. (2007). Video quality metric for bit rate control via joint adjustment of quantization and frame rate. *IEEE Transactions on Broadcasting*, 53(1), 441–446.
53. Kim, C. S., Jin, S. H., Seo, D. J., & Ro, Y. M. (2008). Measuring video quality on full scalability of H.264/AVC scalable video coding. *IEICE Transactions on Communications*, E91-B(5), 1269–1278.
54. Sohn, H., Yoo, H., De Neve, W., Kim, C. S., & Ro, Y. M. (2010). Full reference video quality metric for fully scalable and mobile SVC content. *IEEE Transactions on Broadcasting*, 56(3), 269–280.
55. Ma, Z., Xu, M., Ou, Y.-F., & Wang, Y. (2012). Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(5), 671–682.
56. Ou, Y.-F., Ma, Z., Liu, T., & Wang, Y. (2011). Perceptual quality assessment of video considering both frame rate and quantization artifacts. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3), 286–298.
57. Ou, Y.-F., Xue, Y., & Wang, Y. (2014). Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions. *IEEE Transactions on Image Processing*, 23(6), 2473–2486.
58. Hu, H., Zhu, X., Wang, Y., Pan, R., Zhu, J., & Bonomi, F. (2013). Proxy-based multi-stream scalable video adaptation over wireless networks using subjective quality and rate models. *IEEE Transactions on Multimedia*, 15(7), 1638–1652.
59. Moorthy, A. K., Seshadrinathan, K., Soundararajan, R., & Bovik, A. C. (2010). Wireless video quality assessment: A study of subjective scores and objective algorithms. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(4), 513–516.
60. Raake, A., Garcia, M. N., Moller, S., Berger, J., Kling, F., List, P., et al. (2008). T-V-model: Parameter-based prediction of IPTV quality. In *Proceedings on IEEE international conference on acoustics, speech, and signal processing*.
61. Garcia, M. N., Raake, A., & List, P. (2008). Towards content-related features for parametric video quality prediction of IPTV services. In *Proceedings on IEEE international conference on acoustics, speech, and signal processing* (pp. 757–760).
62. Garcia, M.-N., Schleicher, R., & Raake, A. (2010). Towards a content-based parametric video quality model for IPTV. In *Proceedings of 3rd international workshop on perceptual quality of systems* (pp. 20–25).
63. Argyropoulos, S., Raake, A., Garcia, M.-N., & List, P. (2011). No-reference bit stream model for video quality assessment of H.264/AVC video based on packet loss visibility. In *Proceedings on IEEE international conference on acoustics, speech, and signal processing* (pp. 1169–1172).
64. Yamagishi, K., & Hayashi, T. (2006). Opinion model for estimating video quality of videophone services. In *Proceedings on IEEE global communications conference* (pp. 1–5).
65. Hayashi, T., Yamagishi, K., Tominaga, T., & Takahashi, A. (2007). Multimedia quality integration function for videophone services. In *Proceedings on IEEE global communications conference* (pp. 2735–2739).
66. Belmudez, B., & Möller, S. (2010). Extension of the G.1070 video quality function for the MPEG2 video codec. In *Proceedings of 2nd international workshop on quality of multimedia experience*.
67. Yamagishi, K., & Hayashi, T. (2008). Parametric packet-layer model for monitoring video quality of IPTV services. In *Proceedings on IEEE international conference on communications*.
68. You, F., Zhang, W., & Xiao, J. (2009). Packet loss pattern and parametric video quality model for IPTV. In *Proceedings on IEEE/ACIS international conference on computer and information science*.
69. Joskowicz, J., Sotelo, R., & Ardao, J. C. L. (2013). Towards a general parametric model for perceptual video quality estimation. *IEEE Transactions on Broadcasting*, 59(4), 569–579.
70. Khan, A., Sun, L., Jammeh, E., & Ifeachor, E. (2010). Quality of experience-driven adaptation scheme for video applications over wireless networks. *IET Communications*, 4(11), 1337–1347.
71. Khan, A., Sun, L., & Ifeachor, E. (2012). QoE prediction model and its application in video quality adaptation over UMTS networks. *IEEE Transactions on Multimedia*, 14(2), 431–442.
72. Zhang, X., Xu, Y., Hu, H., Liu, Y., Guo, Z., & Wang, Y. (2013). Modeling and analysis of skype video calls: Rate control and video quality. *IEEE Transactions on Multimedia*, 15(6), 1446–1457.
73. Gustafsson, J., Heikkilä, G., & Pettersson, M. (2008). Measuring multimedia quality in mobile networks with an objective

- parametric model. In *Proceedings on IEEE international conference on image processing* (pp. 405–408).
74. Saygili, G., Gurler, C. G., & Tekalp, A. M. (2011). Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming. *IEEE Transactions on Broadcasting*, 57(2), 593–601.
 75. Stelmach, L., Tam, W. J., Meegan, D., & Vincent, A. (2000). Stereo image quality: Effects of mixed spatiotemporal resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(2), 188–193.
 76. Brust, H., Smolic, A., Mueller, K., Tech, G., & Wiegand, T. (2009). Mixed resolution coding of stereoscopic video for Mobile devices. In *IEEE 3DTV conference*.
 77. Tam, W. J. (2007). *Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV*. JVT-W094, San Jose, CA, April 2007.
 78. Aksay, A., Bilen, C., Kurutepe, E., Ozcelebi, T., Akar, G. B., Civanlar, R., et al. (2006). Temporal and spatial scaling for stereoscopic video compression. In *Proceedings on EUSIPCO* (pp. 4–8). Florence, Italy.
 79. Smolic, A., Mueller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G. B., et al. (2007). Coding algorithms for 3DTV—A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11), 1606–1621.
 80. Yan, T., An, P., Shen, L., Zhang, Q., & Zhang, Z. (2009). Rate control algorithm for multi-view video coding based on correlation analysis. In *Symposium on photonics and optoelectronics*, 2009.
 81. Fehn, C., Kauff, P., de Beeck, M. O., Ernst, F., Ijsselstein, W., Pollefeys, M., et al. (2002). An evolutionary and optimized approach on 3D-TV. In *Proceedings on international broadcast conference* (pp. 357–365), Amsterdam, The Netherlands, September 2002.
 82. Fehn, C. (2004). 3D-TV using depth-image-based rendering (DIBR). In *Proceedings on picture coding symposium*, San Francisco, CA, USA.
 83. Yuan, H., Chang, Y., Huo, J., Yang, F., & Lu, Z. (2011). Model-based joint bit allocation between texture videos and depth maps for 3D video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(4), 485–497.
 84. Liu, Y., Huang, Q., Ma, S., Zhao, D., Gao, W., Ci, S., & Tang, H. (2011). A novel rate control technique for multi-view video plus depth based 3D video coding. *IEEE Transactions on Broadcasting*, 57(2), 562–571.
 85. Cheung, G., Velisavljevic, V., & Ortega, A. (2011). On dependent bit allocation for multi-view image coding with depth-image-based rendering. *IEEE Transactions on Image Processing*, 20(11), 3179–3194.
 86. Norkin, A., Aksay, A., Bilen, C., Akar, G., Gotchev, A., & Astola, J. (2006). Schemes for multiple description coding of stereoscopic video. In B. Gunsel, A. Jain, A. M. Tekalp, & B. Sankur (Eds.), *Multimedia content representation, classification and security* (Vol. 4105, pp. 730–737). Berlin: Springer.
 87. Karim, H. A., Hewage, C., Worrall, S., & Kondo, A. (2008). Scalable multiple description video coding for stereoscopic 3D. *IEEE Transactions on Consumer Electronics*, 54(2), 745–752.
 88. Duan, Q., Yan, Y., & Vasilakos, A. V. (2012). A survey on service-oriented network virtualization toward convergence of networking and cloud computing. *IEEE Transactions on Network and Service Management*, 9(4), 373–392.
 89. Rahimi, M. R., Venkatasubramanian, N., Mehrotra, S., & Vasilakos, A. V. (2012). MAPCloud: Mobile applications on an elastic and scalable 2-tier cloud architecture. In *IEEE/ACM international conference on utility and cloud computing* (pp. 83–90).
 90. Rahimi, M. R., Ren, J., Liu, C. H., Vasilakos, A. V., & Venkatasubramanian, N. (2014). Mobile cloud computing: A survey, state of art and future directions. *Mobile Networks and Applications*, 19(2), 133–143.
 91. Rahimi, M. R., Venkatasubramanian, N., & Vasilakos, A. V. (2013). MuSIC: Mobility-aware optimal service allocation in mobile cloud computing. In *IEEE international conference on cloud computing* (pp. 75–82).
 92. Zhou, L., Chao, H.-C., & Vasilakos, A. V. (2011). Joint forensics-scheduling strategy for delay-sensitive multimedia applications over heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, 29(7), 1358–1367.
 93. Li, P., Guo, S., Yu, S., & Vasilakos, A. V. (2012). CodePipe: An opportunistic feeding and routing protocol for reliable multicast with pipelined network coding. In *Proceeding of INFOCOM*, 2012 (pp. 100–108).
 94. Masnick, B., & Wolf, J. (1967). On linear unequal error protection codes. *IEEE Transactions on Information Theory*, 13(4), 600–607.
 95. Pavlushkov, V., Johannesson, R., & Zyablov, V. (2006). Unequal error protection for convolutional codes. *IEEE Transactions on Information Theory*, 52(2), 700–708.
 96. Gong, C., Yue, G., & Wang, X. (2011). Message-wise unequal error protection based on low-density parity-check codes. *IEEE Transactions on Communications*, 59(4), 1019–1030.
 97. Razavi, R., Fleury, M., Altaf, M., Sammak, H., & Ghanbari, M. (2009). H.264 video streaming with data-partitioning and growth codes. In *IEEE international conference on image processing* (pp. 909–912).
 98. Nazir, S., Stankovic, V., & Vukobratovic, D. (2011). Unequal error protection for data partitioned H.264/AVC video streaming with raptor and random linear codes for DVB-H networks. In *IEEE international conference on multimedia and expo (ICME)* (pp. 1–6).
 99. Huo, Y., El-Hajjar, M., & Hanzo, L. (2013). Inter-layer FEC aided unequal error protection for multilayer video transmission in mobile TV. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(9), 1622–1634.
 100. Vukobratovic, D. (2012). Unequal error protection random linear coding strategies for erasure channels. *IEEE Transactions on Communications*, 60(5), 1243–1252.
 101. Zhang, X., Peng, X., Wu, D., Porter, T., & Haywood, R. (2009). A hierarchical unequal packet loss protection scheme for robust H.264/AVC transmission. In *IEEE consumer communications and networking conference* (pp. 1–5).
 102. Nazir, S., Stankovic, V., & Vukobratovic, D. (2011). Expanding Window random linear codes for data partitioned H.264 video transmission over DVB-H network. In *IEEE international conference on image processing* (pp. 2205–2208).
 103. Stockhammer, T., & Bystrom, M. (2004). H.264/AVC data partitioning for mobile video communication. *IEEE International Conference on Image Processing*, 1, 545–548.
 104. Nazir, S., Stanković, V., Andonović, I., & Vukobratović, D. (2012). Application layer systematic network coding for sliced H.264/AVC video streaming. *Advances in Multimedia*, 2012, 1–9.
 105. Ha, H., & Yim, C. (2008). Layer-weighted unequal error protection for scalable video coding extension of H. 264/AVC. *IEEE Transactions on Consumer Electronics*, 54(2), 736–744.
 106. Dung, H., & Vafi, S. (2013). An adaptive unequal error protection based on motion energy of H.264/AVC video frames. In *IEEE conference on wireless communications and networking conference* (pp. 4594–4599).
 107. Hewage, C. T. E. R., & Martini, M. G. (2013). Quality of experience for 3D video streaming. *IEEE Communications Magazine*, 51, 101–107.

108. Su, G.-M., Lai, Y.-L., Kwasinski, A., & Wang, H. (2013). *3D visual communications*. London: Wiley.
109. Zhu, W., Ding, W., Xu, J., Shi, Y., & Yin, B. (2014). Screen content coding based on HEVC framework. *IEEE Transactions on Multimedia*, 16(5), 1316–1326.
110. Sugawara, M., & Masaoka, K. (2013). UHD TV image format for better visual experience. *Proceedings of IEEE*, 101(1), 8–19.
111. Park, D. S., Choi, S. Y., Lee, H. Y., Kim, Y. T., Hong, J. Y., & Kim, C. Y. (2013). A new wide-gamut RGB primary set and efficient color encoding methods for ultrahigh-definition television (UHD TV). *Proceedings of IEEE*, 101(1), 18–30.
112. Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, 30(1), 21.
113. Salmon, R., Borer, T., Pindoria, M., Price, M., & Sheikh, A. (2014). Higher frame rates for television. *SMPTE Motion Imaging Journal*, 123(4), 50–55.
114. Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., & Ward, G. (2010). *High dynamic range imaging: Acquisition, display, and image-based lighting*. Burlington: Morgan Kaufmann.
115. Miller, S., Nezamabadi, M., & Daly, S. (2013). Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal*, 122(4), 52–59.
116. Mantiuk, R., Kim, K. J., Rempel, A. G., & Heidrich, W. (2011). HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics*, 30(4), 40.
117. Santos, M. A., Villalon, J., & Orozco-Barbosa, L. (2012). A novel QoE-aware multicast mechanism for video communications over IEEE 802.11 WLANs. *IEEE Journal on Selected Areas in Communications (JSAC)*, 30(7), 1205–1214.



Guan-Ming Su received the B.S.E degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park, U.S.A. He is currently with Dolby Labs, Sunnyvale, CA. Prior to this he has been with the R&D Department, Qualcomm, Inc., San Diego, CA; ESS Technology, Fremont, CA; and Marvell Semiconductor, Inc., Santa Clara, CA. His research interests

are multimedia communications and multimedia signal processing. He is the inventor of 45 U.S. patents and pending applications. He is the co-author of *3D Visual Communications* (John Wiley and Sons, 2013). Dr. Su served as an associate editor of *Journal of Communications*; guest editor in *Journal of Communications* special issue on *Multimedia Communications, Networking, and Applications*; and Director of review board and R-Letter in *IEEE Multimedia Communications Technical Committee*. He also served as the Publicity Co-Chair of *IEEE GLOBECOM 2010*, *International Liaison Chair in IEEE ICME 2011*, *Technical Program Track Co-Chair in ICCCN 2011*, *TPC Co-Chair in ICNC 2013*, *Chair in Multimedia for Humanity Theme Track in ICME 2013*, *TPC Chair in ICNC 2014*, *Demo Chair in SMC 2014*, *General Chair in ICNC 2015*, and *Area Co-Chair for Multimedia Applications in IEEE ISM 2015*. He is the Executive Director of *Industrial Governance Board in Asia Pacific Signal and Information Processing Association (APSIPA)* since 2014. He is a Senior member of *IEEE*.

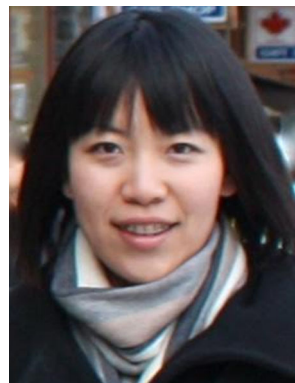


Xiao Su is a Professor and Chair of the Computer Engineering Department at San Jose State University. She received her Ph.D. and M.S. degrees in Computer Science from the University of Illinois at Urbana-Champaign and her B.S. in Computer Science and Engineering from Zhejiang University in China. Her research interests include distributed systems, multimedia communications, network security, cloud computing and big data. She has published over 60 articles in *IEEE* and *ACM* sponsored conferences and journals. She has been a principle or co-principle investigator in multiple grants, from NSF, NASA and industry. She was a recipient of the Charles Davidson College of Engineering Applied Material's award for teaching excellence in 2012, a recipient of Charles Davidson College of Engineering Faculty Award for excellence in scholarship in 2010, and a recipient of NSF CAREER award in 2006.



Yan Bai is an Associate Professor in the Institute of Technology and a founding faculty member of the Information Technology and Systems program at University of Washington Tacoma. She received the Ph.D. in Electrical and Computer Engineering from the University of British Columbia, Canada in 2003. Her research interests include computer networking, multimedia communications, cybersecurity, eHealth, Internet of Things, and cloud

computing. She has published over 50 refereed papers in these areas. She has served as a Technical Program Committee member for over twenty *IEEE* conferences and workshops, and as a reviewer for a wide range of high impact research journals and *ACM/IEEE* flagship conferences. She also serves on the Editorial Board for *IGI Global International Journal of E-Health and Medical Communications*, and the *EAI Transactions on Industrial Networks and Intelligent Systems*.



Mea Wang is currently an Associate Professor in the Department of Computer Science at the University of Calgary. She received her Bachelor of Computer Science (Honours) degree from the Department of Computer Science, University of Manitoba, Canada, in 2002. She received her Master of Applied Science and Ph.D. degrees from the Department of Electrical and Computer Engineering at the University of Toronto, Canada, in 2004 and

2008, respectively. Dr. Wang's research focuses on the design and development of networking systems and algorithms, with special interests in multimedia networking, cloud computing, and peer-to-peer networking.



Athanasios V. Vasilakos is recently Professor with the Lulea University of Technology, Sweden. He served or is serving as an Editor for many technical journals, such as the IEEE Transactions on Network and Service Management; IEEE Transactions on Cloud Computing, IEEE Transactions on Information Forensics and Security, IEEE Transactions on Cybernetics; IEEE Transactions on Nanobioscience; IEEE Transactions on Information

Technology in Biomedicine; ACM Transactions on Autonomous and Adaptive Systems; the IEEE Journal on Selected Areas in communications. He is also General Chair of the European Alliances for Innovation (www.eai.eu).



Haohong Wang As the General Manager of TCL Research America, Dr. Haohong Wang oversees the R&D activities in North America for TCL, a global leader in consumer electronics with over 75,000 employees worldwide and ranks the 3rd in TV and the 5th in mobile phone business. Inspired by the vision of future human-centric smart life enabled by mobile intelligence, he leads his international product and R&D teams in 4 locations to build

cutting-edge technologies in human computer interaction, machine

learning, data mining, mobile sensing, and cloud computing for the next generation consumer products. Before joining TCL, he held technical and management positions at AT&T, Catapult, Qualcomm, Marvell, and Cisco. Dr. Wang is a world-recognized leader in research community as a recipient of the IEEE MMTC Distinguished Service Award, a co-author of five books, and an inventor of over 70 US patents and pending applications. He is the Vice President of Asia-Pacific Signal and Information Processing Association, and chairs the IEEE Technical Committee on Human Perception and Multimedia Computing. He has been the Editor-in-Chief of the Journal of Communications (2010-2014), a Steering Committee member of IEEE Transaction on Multimedia, the Chair of the IEEE Multimedia Communications Technical Committee, the TPC Chair of the IEEE GLOBECOM 2010, and the General Chair of IEEE ICME 2011, IEEE VCIP 2014, IEEE ISM 2015 and ACM Multimedia 2017. He received his Ph.D. from Northwestern University, Evanston, USA.