

Received May 18, 2022, accepted May 25, 2022, date of publication June 6, 2022, date of current version June 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3180491

Measuring, Modeling and Integrating Time-Varying Video Quality in End-to-End Multimedia Service Delivery: A Review and Open Challenges

CHAMINDA T. E. R. HEWAGE¹, (Senior Member, IEEE), ARSLAN AHMAD¹, (Member, IEEE), THANUJA MALLIKARACHCHI¹, (Member, IEEE), NABAJEET BARMAN^{2,3}, (Member, IEEE), AND MARIA G. MARTINI², (Senior Member, IEEE)

¹Cybersecurity and Information Networks Centre (CINC), Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff CF5 2YB, U.K.

²School of Computer Science and Mathematics, Kingston University London, London KT1 2EE, U.K.

³Brightcove U.K. Ltd., London EC4A 3TW, U.K.

Corresponding author: Chaminda T. E. R. Hewage (chewage@cardiffmet.ac.uk)

This work was supported by the Cybersecurity and Information Networks Centre (CINC), Cardiff School of Technologies, Cardiff Metropolitan University, U.K.

ABSTRACT The multimedia delivery chain consists of multiple stages such as content preparation, content delivery via Over-The-Top delivery network and Internet Service Providers network. Within the multimedia service chain, each stage influences the Quality of Experience (QoE) of the end user. The objective of this work is to provide a comprehensive literature survey with future research challenges and opportunities in the field of time-varying video quality in multimedia service delivery. The contribution of this work is two fold: 1) Survey – we provide a review of state-of-the-art works for video quality models to quantify multiple artifacts into a single QoE metric, pooling strategies for global quality measurements, and Continuous Time-Varying Quality (CTVQ) models; 2) Future Challenges and Directions – we investigate ten major research challenges and future directions based on the state-of-the-art for QoE modelling, QoE-aware encoding/decoding and QoE monitoring/management of multimedia streaming in next-generation networks.

INDEX TERMS Time-varying video quality, Quality of Experience (QoE), quality models, video streaming, continuous time-varying quality, video compression, QoE-aware video encoding and decoding, QoE monitoring, QoE-aware network and service management.

I. INTRODUCTION

High resolution multimedia contents and applications are proliferating due to the advancements in consumer electronics, communication and compression technologies. It is expected that over 80% of the Internet data traffic will consist of video data, and over 66% of the connected TV sets will support 4K [1]. This exorbitant increase in video contents in communication networks will demand extortionate bandwidth requirements. To this end, it is envisaged that 10% of the global mobile devices connected to the network will be 5G capable (13.1 billion) by 2023. The enormous growth

The associate editor coordinating the review of this manuscript and approving it for publication was Anton Kos¹.

of multimedia content in communication networks makes it crucial for service providers, content distributors and creators to monitor and measure the Quality of Experience (QoE) for the content transmitted over unreliable and time-varying channels. More importantly, QoE modelling, measuring and monitoring should take place within different stages of the multimedia service delivery chain. Thus, having an in-depth understanding of the elements in end-to-end multimedia service delivery chain, mechanisms for QoE modelling, measuring and monitoring, and QoE-aware operations at each stage in the service delivery chain is important for improving end-user experience in multimedia applications.

To this end, Fig. 1 shows the End-to-End (E2E) multimedia service delivery chain with the development of the

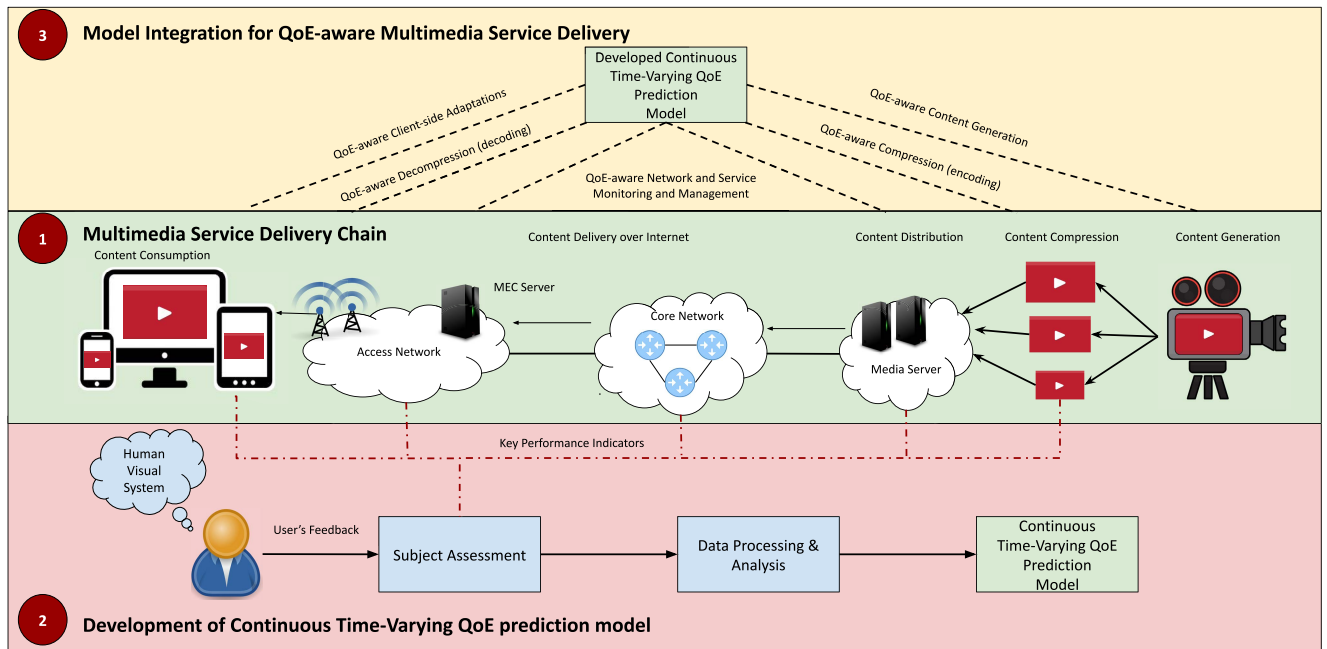


FIGURE 1. An illustration of the 1) End-to-End multimedia service delivery chain, 2) elements and the process involved in QoE model creation and 3) areas where QoE model can be integrated for QoE-aware operations at each stage of the multimedia service delivery chain.

QoE model and model integration in the multimedia service pipeline. As illustrated in Fig. 1, the End-to-End multimedia service delivery chain for video streaming consists of multiple stages including content generation, content compression (encoding), content distribution through media servers, content delivery over the Internet, and content consumption by users using the client-side implementation of the service applications. The process of development and integration of the QoE model into the service delivery chain contains three-phases that include: 1) multimedia service delivery to end-users; 2) development of the QoE model using subjective assessments; and 3) Model integration for QoE-aware multimedia service delivery on different stages of the multimedia service [2]–[5]. To develop a quality (QoE) prediction model for a video streaming service, subjective assessments are performed to collect feedback from multiple users and Key Performance Indicators (KPI) which are further analyzed to develop a QoE prediction model [6]–[9]. Thereafter, the developed QoE model is integrated in several stages in the multimedia service delivery chain [10]–[13]. Such an integration facilitates converting a traditional multimedia service delivery chain into a QoE-aware multimedia service delivery. As illustrated in Fig. 1, this can lead to 1) QoE-aware client-side content adaption, 2) QoE-aware content generation, 3) QoE-aware video encoding and 4) decoding, and 5) QoE-aware network service monitoring and management [14].

A. BACKGROUND AND MOTIVATION

Adaptive streaming has become the norm (as opposed to progressive streaming with a constant bit rate) when delivering multimedia contents to end users through unreliable

and time-varying channels. This is typically achieved through the use of 1) scalable video streams or 2) bitstreams with multiple bit rates coded into layers of contents such that they are adapted according to the network bandwidth available for the end-user. The scalable video streams [15], [16] are generally used with Media Aware Network Elements (MANE) [17] that operate on the server-side (or in edge network nodes). MANEs are capable of adjusting the quality of the video stream by removing or compounding different layers in the bitstream depending on the available network bandwidth [18]. Hypertext Transfer Protocol (HTTP) Adaptive Streaming (HAS) on the other hand operates on the client-side [19] and facilitates adapting the bit rate of the content dynamically to suit the network capacity at a given time [19]. The HAS is now supported by major vendors such as Apple [20], Adobe [21], and Microsoft [22] with their proprietary streaming protocols. Furthermore, the Dynamic Adaptive Streaming over HTTP (DASH) standard introduced by ISO MPEG [19] has made adaptive streaming ubiquitous in the OTT streaming market. In addition, the recent developments in Scalable High Efficiency Video Coding (SHVC) have promoted the use of Scalable Video Coding (SVC) together with HTTP adaptive streaming [23]–[25] to provide further adaptation capabilities.

Even though the existing HAS solutions are promising, maintaining a high consistent end-user QoE is a compelling challenge. For example, the frequency of the dynamic rate adaptation during a media playback has a significant impact on the user experience. Furthermore, associative memory conditions such as primacy, recency, and hysteresis [26] of end-users should be carefully considered when designing effective adaptation algorithms [27]. For example, the impact

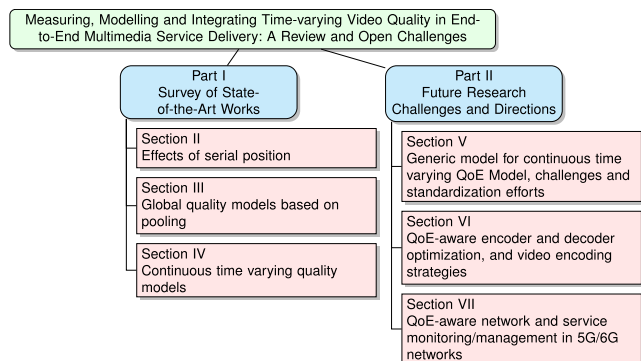


FIGURE 2. Organization and structure of the paper.

of a poor quality video segment will affect user perception of the overall video quality up to a certain duration [28]. In addition, the work in [29] discusses the prospect theory which states that quality loss carries more weight than gain in terms of human visual perception. The awareness of these behaviors of our Human Visual System (HVS) is important when designing efficient rate adaptation algorithms [30], [31].

One of the important elements that enable effective rate adaptation is the ability to accurately predict the time-varying user's QoE. Such prediction algorithms should operate in real-time, account for primacy / recency properties as well as the nonlinear behavior of the HVS [32]. Another factor that impacts end user QoE is the underlying video encoding and decoding algorithms. The modern video coding standards such as High Efficiency Video Coding (HEVC) and Versatile Video Coding (VVC) demonstrate significant coding efficiency improvements over their predecessors. However, the complexity of the algorithms demands excessive use of computational and energy resources at both ends of the multimedia service delivery chain. Hence, certain optimization strategies are carried out both at the encoders and decoders to reduce the energy/compute resource requirements while keeping the user QoE intact. In addition, QoE-aware video compression, bit allocation and rate control, QoE-aware error resilience and concealment, and QoE monitoring and management in communication networks contribute immensely towards maintaining a high end-user QoE in multimedia applications, yet at the same time present compelling research and engineering challenges.

In this context, engineering solutions for measuring modelling and integrating time-varying video quality in E2E multimedia delivery requires an in depth holistic understanding of key parameters that operate in all stages of the multimedia service delivery chain illustrated in the Fig. 1.

B. RELATED WORK

To this end, a number of recent surveys and reviews have been conducted to analyze the state-of-the-art as well as the challenges and opportunities for QoE measuring,

modelling and integration. However, these surveys typically overlook the time-varying aspects of QoE and other stages in the E2E multimedia service delivery chain. These related surveys, their focus areas, and properties that are overlooked compared to this review are summarised in the Table 1. More importantly, as illustrated in Table 1, there is a gap in surveys and reviews that discuss the challenges in QoE management in 5G/6G networks, and the impact on QoE due to the complexity & energy consumption of encoding and decoding algorithms that play a significant role in multimedia service delivery chain. In addition, the application layer techniques such as QoE-aware video encoding, and QoE monitoring in next-generation communication networks have been overlooked in recent surveys and tutorials that discuss QoE modelling.

C. SCOPE AND CONTRIBUTION

Therefore, the objective of this work is to equip the reader with a comprehensive survey of state-of-the-art works and with the future directions in the domain of the time-varying video quality in E2E multimedia service delivery chain. The contribution of the work consists of the following two parts:

- 1) **Survey:** We present a detailed review of existing quality models for quantifying different artifacts into a single QoE metric, pooling strategies employed to obtain global quality measures, and CTVQ models. We also provide the standardization efforts related to time-varying quality estimation.
- 2) **Future challenges and directions:** Based on state-of-the-art works, we provide the ten major future challenges and research directions in the following categories:

- **QoE modelling** – The challenges and opportunities for measuring time-varying video quality modelling and prediction are investigated including 1) Real-time CTVQ prediction – limitations related to real-time utilization of dynamic CTVQ prediction model are discussed; 2) Datasets and Tools – challenges and limitations of available databases/tools are highlighted and future research directions are investigated; 3) Standardization – challenges and future directions associated with the existing standardization activities of HTTP streaming are discussed.
- **QoE-aware encoding/decoding** – We investigate the challenges and research directions related to QoE-aware encoder/decoder optimization that includes: 1) QoE-aware image/video compression – challenges related to effective rate controlling with emerging standards are discussed; 2) QoE-aware encoder optimization – open challenges regarding reduction in encoding complexity, transcoding between different standards and use of machine learning for video coding are investigated; 3) QoE-aware decoder optimization – open research challenges such as embedding decoding complexity data in bit stream and energy consumption are

TABLE 1. Related surveys, their focus areas and limitations.

Related survey	Title	Focus area and factors considered	Limitations and overlooked areas
Barakabitze <i>et al.</i> [12]	QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey	Provides a high-level description of QoE management for multimedia services, which integrates QoE modelling, monitoring, and optimization. Furthermore, authors discuss HAS solutions as the dominant technique for streaming videos over the best-effort Internet.	The authors fail to capture the details and certain aspects of time-varying quality evaluation models which are crucial for continuous QoE monitoring. Moreover, QoE-aware encoding/decoding in multimedia communication is not discussed. Furthermore, future research challenges related to model integration for QoE management are not discussed.
He <i>et al.</i> [33]	A survey of QoE-driven video streaming over cognitive radio networks	Elaborates QoE-driven video streaming over Cognitive Radio (CR) networks	This study does not consider modelling of QoE via CR networks.
Torres Vega <i>et al.</i> [34]	A Review of Predictive Quality of Experience Management in Video Streaming Services	Presents the usage of Machine Learning (ML) techniques for QoE prediction and management	The ML methods analyzed do not focus on time varying quality of a video signal.
Ahmad <i>et al.</i> [35]	Supervised-learning-Based QoE Prediction of Video Streaming in Future Networks: A Tutorial with Comparative Study.	Presents a tutorial on the development and deployment of the QoE measurement and prediction solutions for video streaming services based on supervised learning ML models	The work is limited to QoE prediction/measurement of video streaming services using supervised learning approaches only. The work is only a tutorial followed by comparative analysis that is limited towards only short video sequences.
Zhao <i>et al.</i> [36]	QoE in Video Transmission: A User Experience-Driven Strategy	Discusses the importance of user experience-driven strategy which puts special emphasis on the contextual and human factors in addition to the transmission system	The QoE monitoring and integration aspects are not discussed in detail
Barman <i>et al.</i> [37]	QoE Modeling for HTTP Adaptive Video Streaming—A Survey and Open Challenges	The work provides a survey on QoE models, subjective testing methods and influencing factors in HAS only	The work is limited to the application level KPIs and does not cover the impact and challenges of QoE measurement/management in future networks. Additionally, QoE-aware encoding/decoding is not discussed.
Seufert <i>et al.</i> [4]	A Survey on Quality of Experience of HTTP Adaptive Streaming	The work presents a survey on QoE of HAS applications. The major contribution is toward client-side adaptations and application level influencing factors	The work doesn't provide details related to challenges of QoE monitoring/QoE management in future networks and QoE-aware encoding/decoding.

discussed; 4) Error concealment and resilience-aware video coding – future research direction and challenges associated with intelligent concealment-aware picture partitioning schemes and communication channel-aware encoding strategies are investigated.

- **QoE monitoring and management** – We discuss open research challenges and future directions regarding the integration of the QoE model for QoE-aware network/service monitoring and management in next-generation networks including 1) QoE management in 5G/6G networks – where network enabling technologies can provide an opportunity for deployment and automation of QoE-aware resource optimization in future networks. The challenges related to network/service automation, information exchange among service and network providers, secure and standardized interfaces for QoE KPIs extractions and QoE-aware business model are investigated; 2) QoE Monitoring – active/passive multi-layer QoE monitoring solutions are discussed where issues related to user privacy, extension of existing interfaces/data representation and trade-off between monitoring frequency and control plane traffic are highlighted; 3) QoE model integration & QoE measurement – challenges and future research regarding development and deployment of long-term and more accurate cross layer QoE measurement models are discussed.

D. PAPER STRUCTURE

In this regard, this article is organized as follows: Section II discusses the primacy and recency effects of users during viewing. The global quality models based on different pooling strategies are discussed in Section III. This section also elaborates on different temporal, spatial, and hybrid pooling strategies. Section IV presents the CTVQ models. A generic CTVQ model is introduced in Section V. This section also elaborates the existing challenges in designing CTVQ models and standardization efforts. Sections VI and VII illustrate the challenges in QoE-aware video encoding strategies & encoder, decoder optimization challenges, and QoE-aware network and service monitoring & management in 5G/6G networks, respectively. Fig. 2 represents an overview of the organization and structure of this paper, and a list of abbreviations used in this article is provided in the Table 2.

II. EFFECTS OF SERIAL POSITION OF ARTEFACTS/QUALITY CHANGES

In order to model the QoE for a particular application, user experience over time has to be considered. The effect of memory in assessing the impact of service interruptions, delay, jitter on QoE has been studied for instance in [38]. In general, it has been observed that the same quality values for a frame or portion of a video sequence have a different impact on the global perception of quality depending on the

TABLE 2. List of acronym used in this article.

Abbreviation	Description	Abbreviation	Description	Abbreviation	Description
AV1	AOMedia Video 1	ITU-R	ITU Radiocommunication Sector	RDO	Rate Distortion Optimization
BMA	Boundary Matching Algorithm	ITU-T	ITU Telecommunication Standardization Sector (ITU-T)	ROI	Region of Interest
CAPEX	CAPital Expenditures	JVET	Joint Video Experts Team	RPC	Remote Procedure Calls
CLV	Customer Lifetime Value	KQIs	Key Quality Indicators	RR	Reduced Reference
CNN	Convolutional Neural Network	LMS	Least Mean Squares	RTP	Real Time Transport Protocol
CPU	Central Processing Unit	MANE	Media Aware Network Elements	SAND	Server And Network Assisted DASH
CR	Cognitive Radio	MANO	Management and Orchestration	SDN	Software-Defined Networking
CTU	Coding Tree Unit	MEC	Mobile Edge Computing	SHVC	Scalable High Efficiency Video Coding
CTVQ	Continuous Time Varying Quality	ML	Machine Learning	SLAs	Service Level Agreements
DASH	Dynamic Adaptive Streaming over HTTP	MOS	Mean Opinion Scores	SPS	Sequence Parameter Set
DCT	Discrete Cosine Transform	MOVIE	MOtion-based Video Integrity Evaluation	SROCC	Spearman's Rank Correlation Coefficient
DVFS	Dynamic Voltage and Frequency Scaling	MPEG	Motion Picture Expert Group	SSIM	Structural Similarity
E2E	End-to-End	MSE	Mean Square Error	STM	Short Term Memory
ELAs	Experience Level Agreements	NAL	Network Abstraction Layer	SVC	Scalable Video Coding
FIR	Finite Impulse Response	NFV	Network Function Virtualization	SVM	Support Vector Machines
FMO	Flexible Macroblock Ordering	NR	No-Reference	TCP/IP	Transmission Control Protocol/Internet Protocol
HAS	HTTP Adaptive Streaming	OPEX	Operational Expenditures	TMVP	Temporal Motion Vector Prediction
HD	High Definition	OTT	Over-The-Top	TVSQ	Time Varying Subjective Quality
HDR	High Dynamic Range	PLCC	Pearson Linear Correlation Coefficient	UHD	Ultra High Definition
HEVC	High Efficiency Video Coding	PPS	Picture parameter set	VCL	Video Coding Layer
HM	Harmonic Mean	PSNR	Peak Signal to Noise Ratio	VNFs	Virtual Network Functions
HTTP	Hypertext Transfer Protocol	QA	Quality Assessment	VPS	Video Parameter Set
HVS	Human Visual System	QoE	Quality of Experience	VQA	Video Quality Assessment
IIR	Infinite Impulse Response	QoS	Quality of Service	VQM	Video Quality Metric
IQM	Image Quality Metrics	QP	Quantization Parameter	VQPooling	Video Quality Pooling
ISP	Internet Service Provider	RD	Rate Distortion	VVC	Versatile Video Coding

position at the time of the corresponding frame / portion of the video. This is due to the features of HVS, including recency, primacy, and the asymmetrical response to quality variations of the HVS. These major features contributing to the HVS ([26], [38], [39]) are highlighted and represented in Fig. 3 and the rest of the section discusses each of these key features in details.

A. RECENCY, PRIMACY AND HYSTERESIS EFFECTS

Recency is a characteristic of human subjects and the exploration of recency can reveal the true nature of human perception and behavior. Based on the serial position effect, a viewer tends to remember the last and first items from a series. Previous research [26] suggests that the recall accuracy varies as a function of time. The ability to recall previous experiences which occurred recently is called the “recency effect”, whereas the ability to recall the data from the very beginning of the content is called the “primacy effect”. These effects are said to be caused by the storage of serial information in long-term memory (primacy effect) and working memory (recency effect). The influence of primacy and recency effects on identifying/memorizing certain stimuli is studied in Short Term Memory (STM) research presented in [40]. Studies in particular, in the context of audio and video quality degradations and improvements are reported in [41], [42] and [28], [39], respectively. For instance, the study described in [39] observes that global quality ratings are reduced when a worst-quality video segment occurs at the end sequence compared to the beginning of a 30s video sequence. Furthermore, the authors claimed that the

effect of recency was eliminated when subjects were asked to continuously evaluate picture quality. The duration of the impairment was found to have little impact on quality ratings. A regression analysis carried out in this study also found that quality ratings are best predicted by the peak impairment intensity. These models have been integrated into QoE prediction models which mimic the exponential decay or rise of user QoE over time (e.g., [43], [44]). An audio quality model ITU-R BS.1387 based on recency has been standardized [45]. Not all studies have shown a significant effect of recency. For instance, the time-varying speech quality study described in [42] states that no significant effect is observed during quality switching under the considered experimental conditions. The authors, therefore, concluded that the fundamental human integration operation is not cofounded by speech quality history [42]. Even though these claims depend on the setup and the environment of the tests, similar studies are necessary to clearly understand the true effects of recency for time-varying quality adaptations. The memory of poor quality elements in the past causes subjects to provide lower quality scores immediately afterward, even after the time-varying video quality returns to acceptable levels. This has been denoted as the “hysteresis” effect [32].

B. FREQUENCY OF QUALITY SWITCHING

Due to the increase in multimedia delivery over time-varying channels, audio-visual QoE fluctuations over time have become one of the main focal points in recent research and development in this field. In recent studies, it has been found that temporal fluctuations of media quality taking place

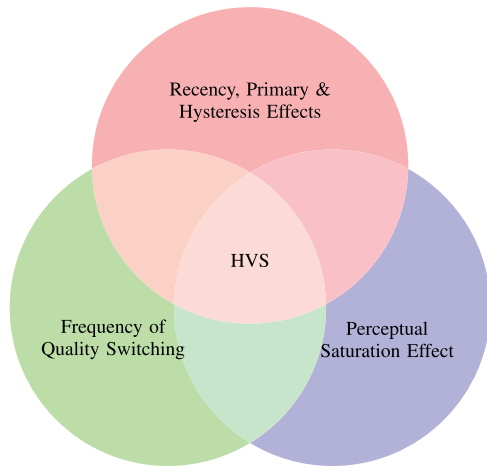


FIGURE 3. Key features contributing to HVS.

within a time scale between 15 secs up to several minutes are governed by the short-term or working memory [41], [46]. The memory effect is linked to the adaptation frequency of the multimedia service. For instance, if the services are frequently adapted (e.g., Adaptive HTTP live streaming enables choosing up to a *second* long video segment) the effect could be much worse compared to the effect when the service degradation/improvement frequency is significantly low. For instance, if the quality of a media stream changes rapidly, the users do not have enough time to settle for a particular viewing condition, which could lead to discomfort due to the high cognitive load associated with switching. In the case of slow media adaptations, the HVS system may get a considerable amount of time to adjust to the current viewing experience. Therefore, it is important to study the memory effect on both fast and slow quality rise/fall situations. Studies in the literature have found some optimum perceptual bounds for the frequency of adaptation for certain applications. For example, perceptual bounds for segmentation sizes for adaptive HTTP streaming are studied in [31], [47].

C. ASYMMETRIC RESPONSE OF HVS TO QUALITY VARIATIONS

The asymmetric nature of users’ responses to quality improvements and degradations are explored and discussed as a mechanism to model the HVS response for time-varying video quality. For instance, users respond more adversely to a drop in quality compared to a quality improvement at a similar degree.

This phenomenon is described by the prospect theory introduced by Kahneman and Tversky in 1979 [29]. Fig. 4 illustrates the dynamics of user response according to the prospect theory. This concept is applied in [30], [31] and [48] to model time-varying quality of rate adaptive video. The detailed subjective studies conducted in these tests show high correlation with Mean Opinion Scores (MOS) for rate adaptive video applications.

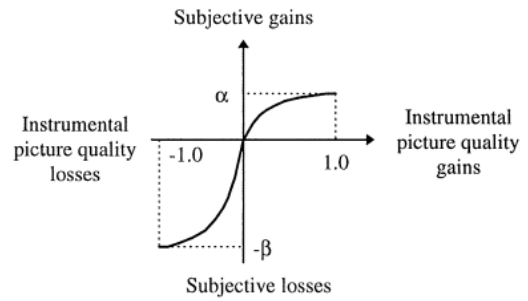


FIGURE 4. Asymmetric response of HVS.

D. PERCEPTUAL SATURATION EFFECT

Perceptual saturation effects also affect time-varying quality. For instance, the quality improvements or degradations after a certain threshold will not be uniquely distinguished by the end-users (due to the masking effect at these extreme quality levels). As illustrated in Fig. 4, after a certain threshold, users won’t be able to differentiate quality changes and their QoE is saturated after these limits. Therefore, when designing quality models, these effects can be taken into account. This will also enable us to fully optimize the usage of system resources. For instance, if the user QoE has passed the saturated level, there won’t be any improvement of user QoE even though we allocate more resources. The model described in [49] applied a predetermined threshold to account for these perceptual saturation effects.

E. SUMMARY

As discussed in this section, it is observed that a number of factors (i.e., hysteresis, the asymmetric response of our HVS, perceptual saturation, etc.) are influencing our HVS when perceiving time-varying quality video. It is paramount to understand these effects and emulate them in CTVQ models. It will be challenging to capture all these perceptual aspects of HVS within a single model. As an alternative, further research can be conducted to evaluate and prioritize these effects, so that new quality models can at least integrate more prominent features.

III. GLOBAL OBJECTIVE QUALITY MODELS BASED ON POOLING

The overall experience or QoE of users is affected by several factors which include capturing artifacts (e.g., lens distortions), compression artifacts, transmission and decoding artifacts. Some of these distortions are spatial artifacts and some are temporal artifacts. The integration of these artifacts into a single quality model is challenging due to their distinguishable and independent characteristics and to how they affect our HVS. Several quality metrics for objective quality assessment exist, such as Mean Square Error (MSE), Structural Similarity (SSIM), Peak Signal to Noise Ratio (PSNR), MOtion-based Video Integrity Evaluation (MOVIE), etc. Most of the metrics, such as MSE, PSNR, and SSIM, were developed for still images and do not

TABLE 3. Objective quality assessment metrics.

Year	Model Type	Metric	Modality
2017	NR	HIGRADE: HDR Image GRADient based Evaluator [51]	HDR images
2017	RR	SpEED-QA: Spatial Efficient Entropic Differencing for Quality Assessment [52]	Video
2016	FR	WMAF: Video Multimethod Assessment Fusion [53]	Images - Frames
2013	NR	NIQE: Naturalness Image Quality Evaluator [54]	Images - Frames
2013	RR	RRED: Reduced Reference Entropic Differencing [55]	Video
2012	NR	BRISQUE: Blind/Referenceless Image Spatial Quality Evaluator [56]	Images - Frames
2011	FR	HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions [57]	HDR images
2010	NR	BIQ: Blind Image Quality Index [58]	Images - Frames
2006	FR	VIF: Visual Information Fidelity [59]	Images - Frames
2004	FR	VQM: Video Quality Metric [60]	Video
2004	FR	SSIM: Structural Similarity Index Metric [61]	Images - Frames
1993	FR	PSNR: Peak Signal-to-Noise Ratio [62]	Images - Frames

necessarily capture temporal distortions in the video, hence the need for sophisticated temporal pooling methods. For example, the SSIM metric developed for images takes into account image degradation, luminance masking and contrast masking. In the MOVIE index, a family of bandpass Gabor filters is used to filter both the reference and distorted videos and the output is then used to measure spatial quality degradation. The output of the spatio-temporal Gabor filters family is then used to calculate the spatial MOVIE index which primarily captures spatial distortions such as blur, ringing, etc. The temporal MOVIE index then captures temporal distortions (e.g., motion compensation mismatch) in the video by tracking video quality along the motion trajectories of the reference video. The spatial and temporal indices are then pooled to obtain a final MOVIE index representing the visual quality of the entire video. Some video quality assessment metrics such as MOVIE correlate quite well with human subjective judgment by evaluating video quality not only in space and time separately but also spatio-temporally. Table 3 provides a summary of the objective quality assessment metrics proposed in the literature and the reader is referred to [50] for a more detailed discussion on the basics of popular terms, metrics, and other related literature.

Pooling can be used to predict the subjective quality of a video using objective metrics by combining the separate effects of spatial and temporal artifacts to obtain a global quality score. Hence a proper choice of the pooling method is crucial for improving the prediction capability of a video quality metric. Pooling can be performed spatially or temporally or in a combined fashion (known as spatio-temporal pooling). Spatial pooling computes the space-varying quality parameters of a video sequence at a single time instant and pools them to obtain a single quality index for that particular time instant. Most of the approaches perform spatial pooling at the frame-level. Temporal pooling then is used to combine these periodical measures over time to get a final measurement for the whole video sequence. Fig. 5

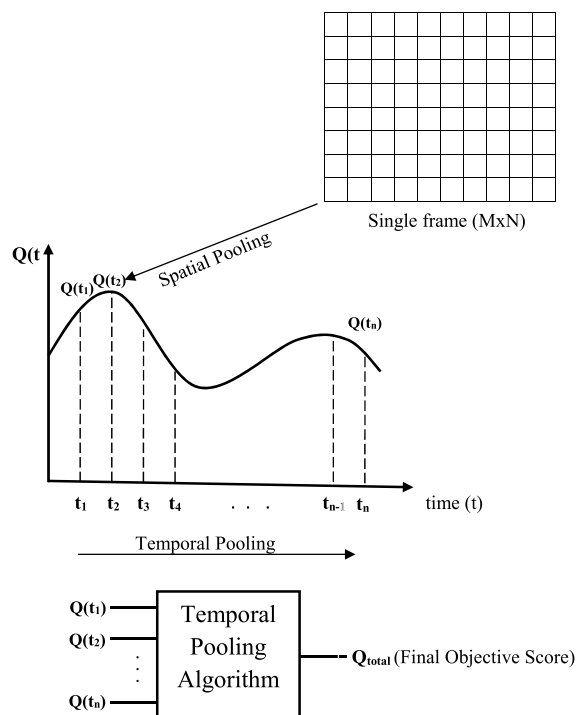


FIGURE 5. Illustration of temporal pooling performed on objective scores obtained by spatial pooling on individual frame.

illustrates a general temporal pooling algorithm to combine periodical quality measures obtained using Spatial Pooling.

A. SPATIAL POOLING

Common spatial pooling methods include simple spatial averaging, Minkowski pooling, Local Quality/Distortion-Weighted Pooling, Information Content-Weighted Pooling [63], Content Adaptive Spatial Pooling [64] and Attention Model (detection of the attention regions in every video frame and average over the distortion map in attention regions only [65]). Spatial pooling methods usually consist of two stages. In the first stage, image quality is evaluated

within local regions (resulting in a quality/distortion map). In the second stage, the spatial pooling algorithm combines the quality/distortion map into a single quality score. For example, Minkowski Pooling for a given quality/distortion map can be defined as $\mathcal{D} = (1/K) \sum_{i=1}^K m_i^p$, where m_i is the quality/distortion value at the i -th spatial location, K is the number of samples and p is the Minkowski power. As p increases it will put more emphasis on image regions with high distortions. A suitable value of p should provide a good approximation of human quality perception. Details on other spatial pooling techniques can be found in [63]–[65].

B. TEMPORAL POOLING

A review of early temporal pooling techniques is presented in [66]. The authors compare the performance of six basic pooling methods applied to five different objective quality metrics using the Pearson Linear Correlation Coefficient (PLCC). Out of the six temporal pooling methods tested in the study (histogram, Minkowski summation, exponentially-weighted Minkowski summation, mean value across a sequence, mean value of scores in the last F frames and local maximum or minimum of mean values of scores in L successive frames), the exponentially-weighted Minkowski summation and the mean value of the last F frames are found to have the best correlation with the subjective ratings. It is observed that the best performing pooling methods are those taking into account the recency effect and the influence of the worst quality section.

All works mentioned so far evaluate the final video quality using local quality information and then combining this temporally, trying to match subjective human scores. In [32] the authors propose a hysteresis based temporal pooling strategy for Quality Assessment (QA) algorithms. In their model they use the average of frame-level quality scores obtained from objective QA algorithms while also taking into account the memory effect of the users (by modelling the quality scores over a certain time duration) as well as the fact that users respond sharply to drops in quality (by sorting the quality scores in ascending order and combining them using a Gaussian weighting function). Some of the latest work on temporal pooling based quality predictions can be found in [67]–[69], and [70].

C. SPATIO-TEMPORAL POOLING

The spatio-temporal approaches allow us to combine both local spatial and temporal image/video features into a single quality metric. The work presented in [77] uses eleven Image Quality Metrics (IQM) for quality assessment on lossy video sequences using temporal pooling methods such as Minkowski summation with different exponents and averaging over distorted frames. Regardless of the IQMs and exponents used, the latter is found to be better than the former. The authors further evaluate different spatial pooling methods such as Minkowski summation over all pixels, averaging over distorted regions and averaging over attention region using average over all the frames as the

temporal pooling method [65]. Furthermore, they evaluate a spatio-temporal scheme using averaging over different distorted spatial regions and frames. Using temporal pooling, spatial pooling and spatio-temporal pooling methods, it is observed that users are more sensitive to distorted spatial regions and temporal segments.

In [75], the authors propose a video quality assessment model which determines the overall quality of a distorted video as a weighted average between global quality and local quality. The global quality is calculated using IQM and direct spatio-temporal averaging method, while the local quality method takes into account visual attention and frequency of quality variations over video frames. Three temporal pooling methods evaluated in this work includes Minkowski summation with exponent 2, direct average of quality values over all frames and a new proposed temporal pooling function which is the filtered result of a function defined as

$$F(k) = \begin{cases} \frac{1}{L}, k \leq \frac{L}{3} & (k : \text{frame index}) \\ \frac{1}{2L}, \frac{L}{3} < k < \frac{2L}{3} & (L : \text{video length}) \\ \frac{3}{2L}, k \geq \frac{2L}{3} \end{cases}$$

by the Gaussian filter for several times (typically 8). The proposed scheme is proved to be better than the other two for all four IQM that were evaluated (PSNR, SSIM, multi-scale SSIM [78] and PSNR-HVS-M [79]).

The model proposed in [76] first calculates the fidelity scores for the video sequence and then pools them to a representative quality score using perceptual motion models. The fidelity measurements discriminate similarity of pixel values between two images. Pooling is performed at frame-level and sequence level using user-defined perceptual weights. The weights are calculated based on the frame type, moving or stationary. The proposed strategy is proved to outperform other VQA algorithms except for MOVIE.

The authors of [64] propose a new content-adaptive pooling strategy called Video Quality Pooling (VQPooling) based on the distribution of local spatio-temporal quality scores from objective VQA algorithms and effects on the perception of large motion, cohesive motion fields such as egomotion (presence of optical flow or motion fields induced by camera motion). Taking into account the slope of the SSIM curve sorted in rank order (steep increasing sections of the curve indicate severely degraded quality section) and presence (or absence) of egomotion, a frame-level quality index is computed in a content adaptive manner. Temporal pooling is then performed by performing low and high quality classification using the K-means clustering method. The scores are then used to calculate the global quality using a weighting function (ratio of low and high quality scores).

The main drawbacks common to the schemes discussed so far are:

- The test video sequences are of very short duration (usually 10-15s)

TABLE 4. Different pooling strategies adopted on time varying quality models.

	Pooling Method	Year	Objective Quality Metrics	Best Correlation Value ^a	Validation Database ^b	Special Remarks
1	Mean Value [74]	2009 ^c	VQM, SSIM, MSE, PSNR, sqrtMSE	-0.8954 (PLCC, VQM)	http://trace.eas.asu.edu/yuv/	Best Prediction Capability: VQM and sqrtMSE; Best pooling methods: Exponential Weighted Minkowski Summation and Mean value of the last F frames.
2	Histogram [74]			-0.9166 (PLCC, VQM)		
3	Minkowski Summation [74]			-0.9159 (PLCC, VQM)		
4	Exponential Weighted Minkowski Summation [74]			-0.9231 (PLCC, VQM)		
5	Mean value of the last F frames [74]			-0.9286 (PLCC, VQM)		
6	Local Maximum (minimum of N frames) [74]			-0.9172 (PLCC, VQM)		
7	VQA balancing local and global quality models [75]	2010	PSNR, SSIM, MSSIM, PSNR-HVS-M	0.945 (PLCC, MSSIM) 0.744 (PLCC, MSSIM)	LIVE EPFL-PoliMI	Performs better than Minkowski Summation and Direct Average
8	Perceptual Motion Model [76]	2010	VQM, PSNR, SSIM, MS-SSIM, MOVIE, T-MOVIE	≈ 0.8 (SROCC) ≈ 0.82 (PLCC)	LIVE	Stable for different distortion types, non-linear in nature, MOVIE is better in some categories
9	Temporal Hysteresis Model [32]	2011	PSNR, SSIM, spatial MOVIE, temporal MOVIE, MOVIE	0.8394 (SROCC, MOVIE) 0.8524 (PLCC, MOVIE)	LIVE	Improved performance compared to other pooling methodologies, takes into account hysteresis effect
10	Video Quality Pooling (content adaptive spatial and temporal pooling) [64]	2013	VSNR, VQM, PSNR, MeanSSIM, SSIM, MOVIE, MSE	0.8427 (SROCC, MOVIE) 0.8611 (PLCC, MOVIE), 0.9335 (SROCC, MOVIE), 0.9422 (PLCC, MOVIE)	LIVE EPFL-PoliMI	Improved performance of most objective quality metrics, recency and hysteresis effect ignored, huge computational burden

^aPLCC: Pearson Linear Correlation Coefficient, SROCC: Spearman’s Rank Correlation Coefficient. The best value along with the metric is represented

^bLIVE Video Quality Database consists of 10 original reference video (9 videos of 10s each and 1 video of 8.68s) and 150 distorted videos [71]; EPFL-PoliMI database consists of 12 original sequences of 10 secs each and 156 distorted sequences [72]; http://trace.eas.asu.edu/yuv/ consist of 4 original video sequences of 12 secs each, 64 progressive CIF video sequences were used in the evaluation [73].

^cYear 2009 indicates the year of study and not necessarily the date the techniques were proposed.

- Only classical approach of video streaming (no quality adaptation over time) is taken into consideration.

All the mentioned temporal pooling schemes so far were evaluated using classical streaming (fixed-bit rate applications) and for short duration videos (typically less than 15s). With the advancement of streaming technologies such as HAS, it is necessary to validate the pooling methods and possibly develop new ones for these latest technologies. The work in [80] evaluates and compares the performance of different temporal pooling mechanisms to validate their application for the recent HAS protocol while also considering long duration video sequences (100s). Based on the results, the authors advocate the use of the simple mean of objective metrics over complicated pooling mechanisms given that the simple mean performs quite well for longer duration sequences. Authors in [81] performed an evaluation of eight different temporal pooling strategies for various objective Video Quality Assessment (VQA) metrics for gaming video streaming applications. Similar to [80], they also observed that no temporal pooling strategy provided a considerable gain over the simple averaging across different VQA metrics. Similar results are also reported by Netflix in [82] where their results suggest that simple arithmetic mean is the best method of averaging per-frame quality scores resulting in a high correlation with the subjective scores. They also observe that Harmonic Mean (HM) produces similar results as simple arithmetic mean but helps to emphasize the impact of small values in the presence of outliers.

D. SUMMARY

This section first describes the objective image and video quality assessment metrics available in the state-of-the-art, then summarised in Table 3. We also highlight the need for global quality models based on pooling. In this regard, state-of-the-art methods based on either spatial or temporal pooling approaches, which yield a global quality value for the whole video segment, are discussed. The strategies discussed above are mainly tested on the LIVE and EPFL-PoliMI databases. Thus, an interesting future work can be to test the performance of all pooling methods for different objective metrics, for both classical and modern streaming technologies, using a single database with various distortions and content types. However, the reported methods (which are summarized in Table 4) are not capable of predicting instantaneous quality values (i.e., CTVQ). In this context, Sec. IV provides an overview of the need and use of Continuous Time-Varying Quality models and their state-of-the-art.

IV. CONTINUOUS TIME-VARYING QUALITY MODELS

This section first discusses the importance and difficulty of modelling continuous QoE in a time-varying image/video streaming application. This allows us to not only anticipate the video’s global quality but also to derive QoE parameters at various time intervals. The presence of such models aids us in making image/video stream adaptation decisions. In adaptive HTTP streaming, for example, a choice is made on which segment should be requested for delivery next. Because of memory effects, it’s best to pick the highest-quality segment

for the next slot. If we can model the time-varying quality with the hysteresis effect and other non-linearities, we can answer this question. Since the inception of video transmission through unreliable communication channels such as wireless channels, the research of time-varying quality has been pursued [28]. The early research was on forecasting overall quality based on quality measurements taken at various points in time. In this scenario, the total quality is predicted using a temporal pooling technique, as explained in Section III [32], [83]. In [32], for example, a temporal pooling method is used to transfer instantaneous objective quality measures to overall video quality using a model that takes into consideration the HVS's recency effect. Even though several of these methods anticipate instantaneous video quality as an intermediate step in calculating overall video quality, the results have not been confirmed against recorded subjective Time-Varying Subjective Quality (TVSQ).

The second generation of continuous quality prediction models can anticipate the rate-adaptive video's instantaneous quality. At the moment, few metrics can model ongoing user responses. The studies in [84]–[87] describe models that incorporate complex low-level vision models. These models are intended to produce a single quality metric for a video stream. Although these algorithms may give frame-by-frame estimates, their temporal summation approaches are not intended to imitate a human subject's continuous quality estimation process. However, the time-varying QoE Indexer proposed by [88] manages to capture interactions between stalling events and then analyze the spatial and temporal content of a video to predict the perceptual video quality. It has also been identified that QoE of an IP transmission can change according to the usage situations. The work presented in [89] suggest that priors related to user attributes can be used to trade-off spatial and temporal quality of an IP stream. Other ways are discussed in [30] and [90]. The method described in [30] combines an objective quality evaluation model at the frame level with a cognitive emulator to account for human viewers' slow temporal responses to image quality changes as well as asymmetric behaviour when picture quality changes from bad to good and vice versa. The distortion masking effect, perceptual saturation effect, and our HVS liking for poor quality experiences over good quality experiences are all modelled in the suggested cognitive model [29]. These methods, however, have drawbacks. Some of these measurements, for example, are designed for slow-changing videos (e.g., video segments lasting 30s or 40s) and so are not ideal for frequent rate modifications [30], [91]. For example, existing HTTP streaming portions could alter in as little as 1-second [92]. Some of the proposed measures are only applicable to low-bitrate videos and do not account for a wide variety of bitrates [85], [92], and [95].

The prospective continuous time-varying quality model predictors, on the other hand, must be simple to deploy in real-time for "on the fly" quality prediction. If the algorithm is too complicated, it will not be able to respond as quickly as needed to adjust the bitrates [91]. As a result, new methods

for TVSQ must be more responsive and comprise basic computations that can be executed on the fly. To consider the hysteresis impact of HVS, the methodologies available in the literature use a variety of approaches. To predict time-varying quality, most techniques use Infinite IIR filter time series. They can often forecast quality, but the complexity of the procedures prevents them from being used in real-time rate adaptation approaches in new technologies like HTTP based streaming. For example, [49] proposes an approach that includes an IIR filter as well as two non-linear filters before and after the IIR filter. To account for the perceptual saturation effects and non-linear response of HVS, two non-linear filters are used at both the input and output stages in this method. If the rate of adaptation is high, this may make a real-time prediction of TVSQ difficult. Hewage and Maria provide a time-varying quality metric based on a moving average filter that accounts for the recency and primacy effects of our HVS in [95]. In comparison to the state of the art, the results of this method indicate good performance. In comparison to other approaches proposed in the literature, the suggested method's easy computations and accuracy allow it to be employed in real-time HTTP based streaming applications. Using a sliding window method, the Cumulative Quality Model (CQM) proposed in [96] predicts the cumulative quality of streaming sessions based on the last window quality, the average window quality, the minimum window quality, and the maximum window quality.

Table 5 summarizes the continuous time-varying quality models reported in the literature to date. The basic building blocks used by the proposed CTVQ models in the literature mimic the hysteresis effect and other HVS responses in addition to the main approach in chronological order. It can be observed that most of the proposed models rely on a dynamic system model and an objective quality model to predict the time-varying quality. This leads the authors to present a generic representation for continuous time-varying quality measurement, which is discussed in Section V.

V. GENERIC MODEL FOR CONTINUOUS TIME-VARYING QoE MODEL, CHALLENGES AND STANDARDIZATION EFFORTS

The analysis of continuous time-varying quality models (in Sec. IV), and pooling-based global objective quality modelling (Sec. III) enabled authors to identify key building blocks for such a model. Based on this, we advocate that a generic CTVQ model should account for spatial/temporal artifacts, perceptual saturation/smoothing and hysteresis effects. Fig. 6 illustrates the proposed conceptual quality model for predicting CTVQ. This model reflects the main concepts (building blocks) employed in the relevant literature discussed in previous sections.

The first module in this model in Fig 6 quantifies the short term quality (i.e., instantaneous video quality) based on both spatial and temporal distortion measures. The objective of this module could be achieved by employing an accurate image/video quality metric that could account for

TABLE 5. Continuous time-varying quality models.

Method	Main Approach	Methodology for calculating hysteresis effect	Other important building blocks
Chen et al [49]	Short term objective quality + Dynamic system model	IIR filter to represent memory effect	Input and output non-linear filters
Seshadrinathan and Bovik [32]	Frame by frame objective quality + time averaging	Gaussian weighting function to assign different weights for previous ratings	-
Tan et al [30]	Objective quality + Asymmetric weighting	Different weights based on prospect theory	-
Masry and Hemami [93]	Objective quality model + dynamic system model	FIR filter to represent hysteresis effect	-
Eswara et al [94]	Linear regression model	Objective short time quality estimates and past TVSQs	
Hewage and Martini [95]	Objective quality model + dynamic system model	Moving average filter to represent hysteresis effect	No pre or post filters

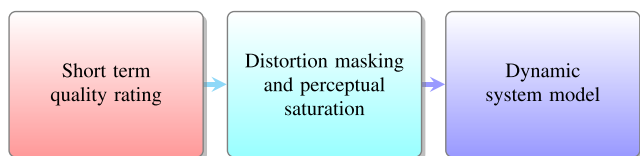


FIGURE 6. Generic continuous time-varying quality model.

both spatial and temporal distortions. This module measures instantaneous video quality in given time intervals. The second module is responsible for the perceptual saturation of the short term quality ratings [49]. Our HVS would not be able to recognize quality gains or drops after a certain threshold (as discussed in Section II.D). Therefore, this module will perform distortion averaging and thresholding to account for perceptual saturation.

Finally, the dynamic system module in the proposed model is responsible to mimic the hysteresis or the memory effect of HVS. In most cases, the dynamic system model could be a Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) filter to perform the temporal decomposition. Pros and cons of using FIR and IIR filters to model hysteresis effects of time-varying video are discussed in [93]. The selection of a particular type of filter could be down to other operational parameters such as delay and complexity. To obtain true user perception under different applications and parameter settings, individual parameters of these components may need fine-tuning.

A. CTVQ MODELLING CHALLENGES

Over the past few years, there have been quite a few models proposed for HAS based applications that tried to predict the continuous time-varying quality and in most cases also the overall QoE score for the whole media session. A review of HAS models provided by the authors in [37] highlights such models. The review paper also highlights various open-source datasets which tried to model such time-varying subjective quality. Also, it is not always possible to obtain short term or long term quality by comparing the processed video with the original content. Especially in real-time applications, the gateway which processes the adaptation may not have access to the original content. In these

scenarios, other approaches such as Reduced Reference (RR) or No-Reference (NR) metrics based on scene characteristics and network characteristics could be used [91], [102]. For instance, the method described in [103] tries to predict quality based on the bitrate. Recently proposed ITU-T Rec. P.1203 also predicts the quality scores on a per-second basis but its performance is not evaluated on a continuous time scale. An interesting future work would be the performance evaluation of such time-varying quality estimation models on a comprehensive dataset.

To build these databases and establish ground truth data, there should be good quality evaluation tools to capture the input of human observers at each time interval. The existing quality evaluation tools for conventional video are neither user friendly nor with enough fidelity to record user inputs at a fast rate. A multimodal interface to obtain 2D/3D quality inputs from the user over time is described in [104]. This tool utilizes external stimuli such as vibration, flickering and sound to improve human concentration during 2D/3D QoE evaluation. The EU FP7 CONCERTO project proposed a comprehensive CTVQ evaluation application interface to be used in mobile phones and tablet computers [105] to measure CTVQ. This tool can be used to evaluate the quality of the stored video as well as with real-time video streaming applications over a wired/wireless network. A snapshot of this APP interface is shown in Figure 7. The application enables to control video playout and the collection of quality scores (from 1 to 5) by every second via a smartphone/tablet. This is done via a slider positioned horizontally on the device’s display. A plot of the provided scores over time is visualised above on the display. A Psychology-based subjective interface is provided by the Bampis *et al.* in [106] which can generate and record visual stimuli with high precision for the collection of continuous, per-frame subjective data. Further research and development should be undertaken to design similar and effective CTVQ measurement tools as mentioned above to duly capture human visual response in future efforts on quality/bitrate adaptation over time.

B. STANDARDIZATION

There are many standardization activities taking place around adaptive bitrate video streaming. For HTTP Adaptive

TABLE 6. Databases for measuring time-varying image quality.

Name of the Database	Video Resolution	Frame Rate	Stalling	Quality switches	Duration	Prediction Scores
Live QoE Database for HTTP-based Video Streaming [50]	720p	30	No	Yes	300 secs	CTVQ + Overall
Waterloo QoE Database [97]	1080p	24,25,30,50	Yes	Yes	10 secs	Overall (No CTVQ)
Live Netflix QoE Video Database [98]	1080p	24,25,30	Yes	Yes	> 1 min	CTVQ + Overall
LFOVIA Video QoE Database [69]	2160p	24,25,30,60	Yes	Yes	120 secs	CTVQ + Overall
Live Mobile Stall Video Database II [101]	1024x576	30	Yes	No	29-134 secs	CTVQ + Overall
Waterloo QoE Database (ECT) [100]	up to 1080p	5,10,30	No	Yes	4,8 secs	CTVQ (4s segments) + Overall
Live NFLX-II Subjective Video QoE Database [101]	1080p	up to 60	Yes	Yes	35 secs	CTVQ + Overall

TABLE 7. Summary of research challenges and future direction in measuring CTVQ.

Research Topic	Research Challenges and Future Research Direction
Real-time CTVQ prediction	Even though a number of dynamic system models are proposed in the literature to predict CTVQ, the application of these models in the end-to-end multimedia chain is limited. The deployment and operational issues (time complexity and real-time resource requirements) are impeding the integration. Furthermore, with regard to learning based QoE modelling, training and execution times have been identified as main challenges. Therefore, further research and development should be carried out to utilize/integrate proposed dynamic system models in real-time applications to maximize end user QoE
Datasets and tools	The availability of open and comprehensive time varying video quality databases are limited for the research community. Only a limited number of public datasets contain instantaneous subjective image quality measurements. The majority contains overall image quality for the whole video (summative image quality). This is a challenge given different types of video content (e.g. SD, HD, 4K, UHD, 3D, etc.). Therefore, further efforts should be made to create good video quality databases with instantaneous subjective image quality measurements or QoE as well as the overall QoE for the whole video. Furthermore, the tools and applications for conducting, measuring, visualizing, analysing and storing subjective quality ratings at different time intervals during subjective quality measurement experiments are rare as discussed in this section. This is a challenge due to the numerous factors that need to be measured over time to understand the QoE as perceived by the end users. This needs to be overcome in order to facilitate efficient subjective quality experiments in the CTVQ domain.
Standardization efforts	Standardization of technologies related to adaptive HTTP streaming is required to provide conformity among different vendors. Further efforts should be made in standardizing subjective quality evaluation of time varying video applications. This will facilitate the standardization of conducting subjective quality measurement experiments in this domain. Furthermore, standardization work on networks domain will support efficient integration of emerging QoE models in the end-to-end delivery chain



FIGURE 7. A snapshot of the authors' mobile APP for evaluating CTVQ.

streaming DASH is the main standardization effort to date [19]. MPEG-DASH was developed by MPEG and became an International Standard in November 2011. Since then MPEG-DASH has been revised as MPEG-DASH ISO/IEC 23009-1:2019 in 2019. This standard defines a media representation framework for dynamic adaptation of media content in an encoder agnostic manner. However, this standard does not address the QoE or time-varying quality of the video. In 2017, Server and Network assisted DASH

(SAND) was published as an extension of the MPEG-DASH standard [107]. This standard focuses on content-awareness and QoE-service-awareness through server/network assistance, analytics and monitoring of DASH-based services, and unidirectional/bidirectional, point-to-point/multipoint communication with and without a session (management) between servers/CDNs and DASH clients. It is observed that even though there are standardization efforts on media representation and reference decoders, no standardization effort is on monitoring and measuring the quality of video delivered using adaptive streaming. The standards such as ITU-R Recommendation ITU-R BT.500-14 [108] also required to revisit to measure the quality of time-varying video quality. For instance, ITU-R Recommendation ITU-R BT.500-14 recommends video segments to be 10s long or use compositions of segments to make a longer video for subjective quality evaluation tests. Similarly, there is a clear lack of open reference video datasets for conducting TVSQ evaluations. In addition, there are no standard tools or procedures to evaluate subjective opinions for time-varying quality videos. Often custom made tools are used to evaluate the quality (e.g. Joystick). These limitations hinder the progress of research in this domain. There are also

TABLE 8. Summary of standardization in HTTP adaptive video streaming.

Standards	Standardization Body	Purpose
MPEG-DASH [19]	ISO	For Content delivery in HTTP adaptive video streaming
SAND [107]	ISO	For server and network-aware content delivery in HTTP adaptive video streaming
BT.500-14 [108]	ITU-T	For measuring time varying video quality
P.1203 [109]	ITU-T	For measuring QoE of HD HTTP adaptive video streaming
P.1204 [110]	ITU-T	For measure QoE of 4K HTTP adaptive video streaming

standardization activities in study group 12 at ITU-T such as ITU-T P.1203 for HD HTTP adaptive video streaming [109] and ITU-T P.1204 for 4K video streaming [110]. The ITU-T P.1203 is a parametric bit-stream model with four different modes of operation where each model utilizes different application layer Key Quality Indicators (KQIs). The works in [111], [112] provided an open-source Python language-based implementation of ITU-T P.1203. Table 8 shows the summary of the standardization efforts in the domain of HTTP adaptive video streaming.

C. SUMMARY

As discussed in this section, even though the basic building blocks for the generic continuous time-varying QoE model can be envisaged, there are many challenges to be addressed in the future. Some of the key challenges and future research directions are discussed in Table 7. It is also important to deploy these CTVQ models in the end-to-end multimedia QoE pipeline that spans from encoding to delivery. Therefore, it is important to discuss the methodologies followed to integrate these QoE models with the networking infrastructure and the challenges associated with QoE monitoring and management. Similarly, the distortions introduced at the video compression, and optimizations at the encoder and decoder have a significant impact on the overall end-user QoE. Hence, QoE modelling should consider the impact of compression and other application-layer processes. In this regard, Section VI elaborates on compression artefacts, the impact of the encoder, decoder optimizations on the QoE and QoE-aware encoding strategies. To this end, Section VII discusses QoE model integration and monitoring and management challenges in 5G/6G. The following two sections combine the CTVQ modelling and deployment within content creation and communication infrastructure in end-to-end multimedia delivery chain illustrated in Fig. 1.

VI. QoE-AWARE ENCODER AND DECODER OPTIMIZATION, AND VIDEO ENCODING STRATEGIES

The QoE modelling aspects discussed in Sections III-V can be applied within the communication infrastructure

(discussed in Sec. VII as well as within the content creation stages (as defined in Fig. 1). For instance, the knowledge and understanding of end-user QoE at the content generation influence the video compression and multimedia adaptation techniques. As such, the QoE-aware multimedia adaptation techniques can be broadly categorized into application layer adaptations and methods that operate on the network layer. Network layer-based techniques constitute a range of sophisticated approaches such as network resource allocation [113], and QoE management methods using SDNs, and NFV, which are discussed in Section VII. Application layer methodologies can also be categorized into adaptations within the media distribution chain, encoder level, and decoder (and media playback device's) adaptations. Multimedia content adaptation within the media service delivery chain typically focuses on selecting the most suitable video segment with a bit rate that matches the prevailing network bandwidth. These techniques discussed in [114]–[116] work well with HTTP adaptive media streaming technologies such as MPEG-DASH to improve the end-user's QoE in fluctuating network conditions [70]. However, these methods do not consider the algorithms that operate at the encoder and decoder level and their impact on the end-user QoE. These aspects are of prime importance when designing theory, technologies, and applications for 5G multimedia communications [117]. Compressing the raw image and video contents leads to compression artifacts and visual quality degradation. This is evident, especially in low bit rate transmission cases. As the number of bits per frame decreases, both the objective and subjective quality of the image sequence will decrease. On the other hand, another major challenge when working with modern video coding standards is their exorbitant increase in the encoding and decoding time complexity. Yet, the adaptation techniques that operate on the distribution pipeline do not address the QoE implications of the optimization strategies undertaken at the modern encoder and decoder implementations. This section provides an overview of the encoding and decoding complexity of modern video coding standards, optimization strategies proposed in the literature, and their impact on the end-user QoE. Furthermore, QoE-aware encoding techniques proposed in the literature to generate bitstreams that can potentially improve end-user QoE in lossy communication channels are also discussed.

A. IMPACT OF IMAGE/VIDEO COMPRESSION ON QoE

It is estimated that over 80% of the internet traffic will consist of video data by 2023 [118] and efficient video compression and communication strategies must be continued to be investigated. The challenge for any video and image encoding algorithm is to compress the visual data while maintaining a target quality level [119]. Modern hybrid video coding standards achieve this by reducing the redundancies seen in the video signals: namely the spatial, temporal redundancies and entropy coding to reduce the redundancies between data symbols (achieved through variable length coding) [120]. The

recent video coding standards are proving to be more efficient compared to their predecessors. For instance, HEVC which was standardized in 2013 shows 40-50% coding efficiency improvement compared to H.264 [121], [122]. The demand for coding efficiencies beyond HEVC was immediately felt with the exponential increase in UHD, HDR video contents, and immersive media applications [123]. Thus, the collaboration of the Joint Video Experts Team (JVET) composed of ITU-T and MPEG members introduced Versatile Video Coding (VVC), in 2020 [124] which demonstrates 30-40% bit rate reduction for the same quality level compared to HEVC. In addition, the popularity of open-source, royalty-free codecs such as VP9 [125] and AV1 [126] has seen an increase in the recent past. Experimental data between different standards reveal comparable compression efficiency performance between VP9/AV1 and their MPEG counterparts. In any case, the aforementioned standards are all based on the hybrid block-based video encoding approach. Hence, the Discrete Cosine Transform (DCT) based compression in block-based video coding achieved through quantization of transformed coefficients generally results in compression artifacts such as blurring, ringing distortions, and boundary artifacts (blockiness) [127]. The degree of visibility of these artifacts affects the end-user's QoE. Hence, reducing the bit rate to cater to the increasing demands in video communications without significant impacts on the perceived visual quality [128] is a compelling research challenge that will continue to dominate this space for the foreseeable future.

Under this umbrella of research, Region of Interest (ROI) based video coding and efficient rate controlling algorithms are seeing an interest in the research community. The proliferation of machine learning algorithms and emerging low complex neural network architectures are facilitating new research avenues over the traditional rule-based algorithms. Rate controlling deals with a complex problem especially in real-time communication and any inaccurate bit rate estimation or allocation coupled with a sub-optimal coding parameter selection can lead to unacceptable QoE issues in multimedia applications. The λ domain rate control algorithm proposed in [129] and adopted in the HM Test Model reference encoder [130] is attempting to dynamically model the RD relationship using a Least Mean Squares (LMS) based adaptation model. A rate controlling algorithm requires 1) allocation of certain amount bits to a frame or a CTU based on the overall bit rate and 2) selecting the appropriate coding parameters to meet the bit rate constraint set at the beginning of the encoding process. The dynamic nature of the content, and the fact that bit allocation takes place at an early stage of the encoding pipeline makes it difficult for the encoders to simultaneously achieve both tasks. Hence, advanced CTU level and frame-level bit allocation methods have been proposed in the recent literature using a range of techniques. Game theory-based methods are presented in [131]–[133] targeting video conferencing applications. If the rate estimation is incorrect for a particular frame or

a CTU, the overall rate control performance is degraded impacting the end-user QoE. Hence, the adoption of machine learning models, Bayesian estimation methods is prominent in recent rate controlling algorithms [134]. The use of Convolutional Neural Network (CNN) for extracting spatial and temporal saliency features, and perceptual priority-based QP selection algorithms are becoming popular within the video coding domain [135]. Identifying ROI areas through a saliency map is important in this regard, as it facilitates the bit allocation algorithms to allocate more bits into these regions of the frame compared to static regions. Deep CNNs are utilized to generate saliency maps to identify ROIs in a frame in [136], [137] and the algorithm is then extended to introduce a modified RDO process to perform an optimum bit allocation based on the importance of the region. Machine learning-based ROI extraction is proven to be more consistent with the HVS compared to the traditional texture based extraction techniques. This is attributed to the fact that these prediction models are specifically trained using large datasets to recognize areas with objects as regions of interests within a video frame. The popularity of machine learning technologies has given rise to a range of deep learning-based image and video coding solutions where the encoder and decoder are composed of neural networks. It has been identified that learning-based image coding solutions offer significant compression performance, yet the decoded images are affected by compression artifacts that are typically not seen with DCT-based coding methods [138]. Therefore, further investigations are needed in assessing the overall QoE impact of compression algorithms, rate controlling methods, and ROI-based coding techniques with state-of-the-art video coding standards.

B. IMPACT OF THE ENCODER OPTIMIZATION ON QoE

Modern video coding standards such as High Efficiency Video Coding (HEVC) [122], Versatile Video Coding (VVC) [139], VP9 [125], and AV1 [140] have shown significant coding efficiency improvements compared to their predominant predecessors such as H.264, and VP8 video coding standards. For instance, HEVC which was introduced in 2013 shows 50% [141] coding efficiency improvement compared to its predecessor H.264. Similarly, VVC which was standardized in 2020 [142] demonstrates 30-35% [143], [144] bit rate reduction compared to HEVC for a similar video quality level. However, these improvements in the bit rate reduction arrive with a greater increase in the encoding complexity which demands a significant amount of computational and energy resources at the encoding servers/devices [144]–[146]. For example, the use of larger coding blocks (64×64 in HEVC, and 128×128 in VVC), increased intra- and inter-prediction modes, cross-component prediction, advanced motion compensation and estimation methods, and improvements in transform coding have resulted in a ≈ 250 -400% encoding time increase compared to its predecessor HEVC [144]. In this context, a significant amount of research effort is dedicated to optimizing the

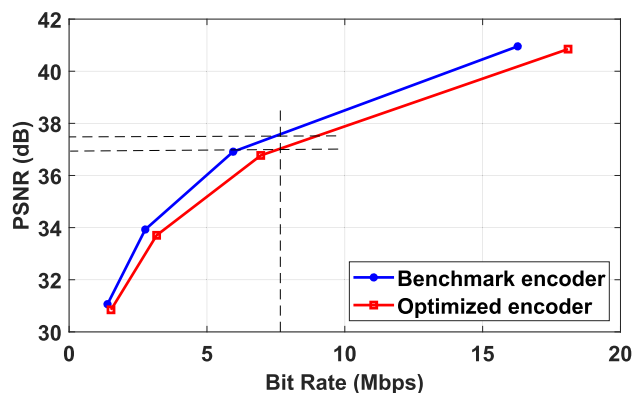


FIGURE 8. An illustration of the RD curve in a typical encoder. The “Benchmark encoder” curve represents the coding performance of an encoder that achieves the best coding efficiency performance. The “Optimized encoder” curve represents the RD performance of an encoder that skips the computationally intensive RD optimization to select the coding modes for a given content.

encoders to reduce the encoding time complexity with minimal impact on the coding efficiency.

Encoders typically follow a Rate-Distortion (RD) optimization using a Lagrangian cost function and go through all possible encoding parameter combinations in a brute force fashion to determine the optimal encoder parameter combination (that minimizes the RD cost) for a given content [147]. Therefore, state-of-the-art methods that focus on encoder complexity reduction attempt to skip all or certain stages in the compute intensive RD optimization when determining the best possible encoding parameter combinations for a given content. In this regard, the majority of the encoder complexity reduction methods can be categorized into two approaches; statistical feature-based methods and learning-based approaches [148]. Statistical feature-based methods attempt to use texture complexity [146], [149], motion complexity details [150], [151], combined with statistics from previously encoded blocks [152] to generate probabilistic models (e.g., Naive-Bayes) to early determine the best coding structure/parameters for a given image segment. On the other hand, learning-based approaches use data sets composed of previously encoded image segments and associated encoding parameters to train machine learning models which can then be used to predict the coding structures/parameters for a given arbitrary content. Supervised learning algorithms such as Support Vector Machines (SVM) have been used predominantly in recent literature due to their less complexity and ability to handle binary classification effectively [148], [153]. In addition, techniques such as random forests [154], decision trees for data mining [155], and various deep learning-based methods [156], [157] have been attempted to predict coding parameters at various stages in the encoding toolchain.

The main challenge that presents with any encoder optimization algorithm is to achieve a significant encoding complexity reduction while maintaining the coding efficiency achieved by the benchmark encoding algorithms presented

in reference implementations [130], [158]. Any optimization strategy that attempts to reduce the encoding complexity by skipping the exhaustive RD optimization tends to select a certain number of less optimum coding parameters for a given image segment [146], [150], [155]. These less optimum selections cause the encoder to operate slightly below the benchmark RD curve as illustrated in Fig. 8. As a result, a bitstream generated at a particular rate by such an encoder will have a lesser objective (as well as subjective) quality level. In such cases, the encoder will have to generate a bitstream at a higher bit rate to achieve a similar visual quality (Fig. 8). These scenarios either result in low quality video streams or will demand a higher network bandwidth that ultimately impacts the end-user QoE. Hence, some of the proposed encoder optimization algorithms tend to provide engineering design parameters to trade-off the coding efficiency to the encoding complexity depending on the end-user’s QoE requirements, network conditions, and computational resource constraints of the encoding servers [148], [150].

C. IMPACT OF DECODER OPTIMIZATIONS ON THE QoE

The assortment of new coding tools and features in novel video coding standards increase the complexity of the resulting bitstreams. Complex bitstreams increase the time complexity of the decoders and demand more computational and energy resources to achieve real-time decoding for smooth video playback. For example, decoding time for VVC bitstreams has increased by ≈ 130 -170% compared to HEVC bitstreams [144] making decoding of these bitstreams a major source of energy consumption in resource constrained consumer electronic devices. This is further compounded by the ever-increasing resolution of the video frames (e.g., HD, 4K, and 8K), and novel media formats such as High Dynamic Range (HDR), 360 degree videos, etc. [159], [160]. Thus, the overall energy consumption at the decoder (particularly in the case of mobile devices), has become a vital parameter that affects the overall end-user QoE in multimedia applications [161], [162].

Reducing the energy consumption of resource constrained consumer electronic devices during media playback is a compelling research and engineering challenge. This is typically attended across all layers of the TCIP/IP stack (physical layer, link layer, network layer, and application layer) [163], [164]. Physical layer approaches mainly focus on the changes to the modulation scheme or dynamic modulation scaling techniques [165], whereas the link layer and network layer techniques operate on managing the wireless interface and energy-aware scheduling algorithms [166]–[168]. The application layer techniques on the other hand use a range of approaches to manipulate the decoder operations (changes to the loop filtering and motion compensation) [169], change the video stream (by manipulating coding parameters such as resolution, frame rate, and Quantization Parameter (QP)) [170], and also by considering decoding complexity-aware video coding at the content preparation

stage at the encoders [159]. It has been identified that changes to the video resolution, and frame rate can result in severe perceived video quality degradation in mobile video broadcasting [171]. However, in the case of the latter, encoders can be designed to consider the corresponding decoding complexity (or energy consumption) associated with a particular coding mode, or coding block size for a given content. For example, the decoding energy parameter is considered within the RD optimization to jointly select the coding modes that attain a minimum distortion for a given bit rate, and decoding energy constraint. This is particularly important as any attempt to reduce the decoding energy consumption results in impacting the coding efficiency and is eventually impacting the QoE. Furthermore, the use of Dynamic Voltage and Frequency Scaling (DVFS) in both software and hardware decoders is evident in modern mobile devices to manage the device's power consumption. For example, the CPU's operating frequency and voltage level are adjusted depending on the complexity of the bitstream [172], [173].

Application layer methods that alter the decoding process to reduce the energy consumption typically skip certain operations within the decoding process. For example, Green-MPEG (a standardization effort from MPEG to reduce the energy consumption of the decoders) proposes to send metadata specifying decoding-complexity requirements for a video frame (or a video segment) [174], [175]. This allows the decoders to skip certain high complex steps in the decoding pipeline to reduce the overall energy consumption [176]. However, it is reported that such alterations (especially within the motion compensation phase) severely affect the video quality, hence the QoE [159].

DVFS on the other hand attempts to reduce the decoding energy consumption by altering the operating CPU frequency and voltage levels depending on the complexity of the current video frame [172], [177]. The main challenge with DVFS is to estimate the decoding complexity of subsequent video frames to set the appropriate CPU frequency level. Inaccurate estimates lead to frame drops, and sub-optimal CPU frequency/voltage levels adversely affect the overall system performance degrading the end-user's QoE in multimedia consumption. The emerging Green-MPEG specifications such as C-DVFS attempts to mitigate these challenges by incorporating decoding complexity/energy requirement metadata into the bitstreams to assist accurate frequency/voltage scaling [178].

D. ERROR RESILIENCE AND CONCEALMENT-AWARE VIDEO ENCODING

Video transmission over a lossy channel is a compelling challenge and requires cross-layer approaches to protect the media streams against network vulnerabilities. In this context, QoE-aware media protection strategies such as forward error correction should be supported by both the encoder and decoder that operate in the application layer to recover the lost information [179]. The authors in [180] classify

the transmission challenges on DCT compressed images as bit-error, desynchronization, packet loss, packet delay, and packet intrusion. In these cases, the encoder incorporates additional information into the bitstream to support error resiliency whereas the decoder utilizes a range of error concealment strategies to recover or conceal the lost data from the reconstructed video frame. The frame or slice copy is the simplest error concealment strategy supported by the majority of the video decoders [181]. However, reconstructing a video frame with a lost video slice in a static, low textured background section of the frame has a low impact on the QoE, compared to a case where the lost video frame slice is in a highly textured, motion rich section of a video frame. In the case of latter, the degradation of the visual quality is easily noticeable to the users and has a high impact on the end-user QoE [182]. As a result, strategies such as Boundary Matching Algorithm (BMA) are adopted to estimate the motion vectors for the pixels of the lost slice from the motion information available in the neighboring slices [183]–[185]. Recovery of full lost frames through texture analysis and motion vector extrapolation is also attempted as decoder-side error concealment techniques [186]. However, handling and reducing the error propagation remains a crucial challenge.

The focus has also been directed towards priority-based slice protection schemes to improve the error correction capability of the decoders. For example, non-Video Coding Layer (VCL) Network Abstraction Layer (NAL) units such as Video Parameter Set (VPS), Sequence Parameter Set (SPS), and Picture Parameter Set (PPS) NAL units are given additional protection within the lower layers of the TCP/IP [187], [188]. One of the prominent application layer strategies to obstruct error propagation in the decoders is the use of intra-predicted blocks or intra-coded frames within the encoded bitstreams. Intra-coding involves predicting the current coding block using the previously encoded spatially adjacent pixels as opposed to the inter-coded frames that predict from previously encoded temporally adjacent video frames. Three approaches have been identified to use intra-coding within a sequence. These include intra coding of several blocks selected randomly or regularly, intra coding of some specific blocks selected by an appropriate cost function, or intra coding of a whole frame. Intra-coding increases the bit rate impacting the compression ratio, but the experimental results demonstrate periodical I-frame coding is preferred over coding only several blocks as intra mode in P-frames [186]. Encoders can configure the number of intra-frames within the Group of Pictures using the configuration parameters such as intra-refresh interval. However, frequent injection of intra-frames increases the bit rate of the encoded stream. Hence, dynamic and intelligent model-based methods to adjust the intra-frame frequency is heavily investigated in the literature [189], [190]. Duplication of macroblocks inside Regions of Interest (ROI) within a video frame introduces additional redundancies yielding high quality reconstruction of frames with lost data packets [191]. These attempts are further extended by Flexible Macroblock

TABLE 9. Research challenges and recommendations in QoE-aware video encoding and decoding.

Research area/topic	Main research challenges and recommendations
QoE-aware image/video compression and streaming	Efficient rate controlling through intelligent bit allocation, and appropriate coding parameter selection are compelling challenges, specially with the new standards such as VVC, AV1. Use of machine learning to identify ROI within a frame is becoming more prominent with these algorithms. Furthermore, standardization activities are ongoing for client-specific manifest files for DASH through Session Based Description (SBD). The demand for efficient compression technologies is rising constantly and emerging MPEG-5 standards are generating opportunities for more efficient and low complex video coding algorithms.
QoE-aware encoder optimizations	Reducing the encoding complexity of state-of-the-art video coding standards (HEVC, VVC, AV1) while keeping the coding efficiency intact is a compelling research challenge. Novel standards such as MPEG-5 LCEVC present further research into low complex video coding and its impact on end-user QoE. Furthermore, transcoding between different standards, use of machine learning for low complex video coding without impacting coding efficiency will require further investigations.
QoE-aware decoder optimizations	The energy profiling of the latest and upcoming video coding standards is crucial. Investigations into embedding decoding complexity data into the bit stream, efficient use of DVFS under different operational conditions, and the consideration of energy consumption of the whole end-to-end multimedia streaming pipeline will be crucial.
Error concealment and resilience-aware video coding	As the coding efficiency improves (with standards such as HEVC, VVC), and redundancies are removed from the source signal, the compressed signal becomes more prone to transmission errors. Intelligent concealment-aware picture partitioning schemes, communication channel-aware encoding strategies need to be further investigated.

Ordering (FMO) and ROI-based rate controlling to improve the error resiliency against bursty packet losses and end-user QoE in lossy channels [192]–[194]. Modern video coding standards such as HEVC and VVC introduces features such as Temporal Motion Vector Prediction (TMVP) to improve their coding efficiency performance [122]. However, TMVP is a major source of error propagation in lossy communication channels and research such as [195], [196] attempt to address this by intelligently turning on and off TMVP depending on the channel conditions and content being encoded.

In this context, QoE-aware encoding parameter selection at the encoder can play a vital role in supporting the decoders in reconstructing the encoded bitstreams such that end-user QoE is maximized. For instance, probabilistic models that predict the overall distortion at the decoder based on the impact on motion vectors, the pixels in the reference frames, and the clipping operations. The predicted error is used in the encoders to select the optimum coding parameters to facilitate robust video transmission in lossy channels [197]. A similar approach is undertaken in [198] where the optimal error concealment strategies at the decoder are identified in the encoder by simulating the transmission errors. These are then signaled to the decoder as supplemental enhancement information messages. Two-pass coding strategies are also considered to determine the motion vectors, coding modes such that both source coding, channel propagation errors are reduced while improving the error concealment capabilities of the decoders [199].

E. SUMMARY

This section describes the challenges and potential research directions in the application layer compression domain techniques for QoE-aware multimedia adaptation, and the key challenges and potential research directions are summarized in Table 9. The QoE impact of the image/video compression and potential strategies to improve the end-user QoE during compression are discussed in Section. VI-A. These include the use of ROI-based bit allocation, and intelligent rate controlling algorithms using state-of-the-art machine

learning models. Section VI-B discusses the impact of the complexity of modern video coding standards on the end-user QoE. The encoder optimization strategies to reduce the complexity impacts the coding efficiency of the encoders. In this case, the resultant bitstreams produce low quality video contents, or will demand higher bit rates. In this context, effectively trading off the complexity to the coding efficiency, and producing bitstreams with low visual quality impact is a compelling challenge. Similarly, the impact of decoder energy optimization strategies are discussed in Section VI-B. In this case, various application layer methods, DVFS strategies and decoding complexity-aware encoding algorithms and their challenges are discussed. The increasing complexity of the encoded streams demands additional computing and energy resources; a scarce resource for mobile handheld consumer electronic devices. Hence, energy consumption during multimedia consumption is a crucial element that affects the overall end-user QoE in future multimedia applications. Section VI-C discusses the error resilient and concealment-aware video coding methodologies available in the state-of-the-art. In this regard, modelling the quality of the reconstructed video sequence of a bitstream transmitted over a lossy channel is important to determine the optimum coding parameters to be used in the encoder. This can assist the encoder to generate a bitstream that is more resilient in the error prone channel and can support decoder’s error concealment strategies in order to maximize the end-user QoE.

VII. QoE-AWARE NETWORK AND SERVICE MONITORING/MANAGEMENT IN 5G/6G NETWORKS

Once the QoE prediction model is developed by following the considerations discussed in Section III-VI, it can be integrated into the multimedia service delivery over the Internet for QoE-aware network and service management/monitoring. The QoE-aware multimedia streaming service delivery in the next-generation networks relies on novel solutions for the integration of the QoE models in the future internet architecture and network and service management chain [12],

[13], [200]–[202]. The effective QoE model integration in the 5G/6G networks for the video streaming requires a QoE monitoring solution to be deployed to monitor Key Quality Indicators (KQIs) based on the features/parameters contributing to the QoE model [161], [203]–[205]. Furthermore, the monitored KQIs are then utilized by the QoE models for the measurement of the user-perceived quality leading to QoE-aware management of the network and service resources [10], [206]–[209]. This section provides the future challenges and research directions with the perspective of the QoE-aware network/service management and monitoring in the 6G and beyond networks by integration of the QoE model in the service chain and future internet architecture.

A. QoE MANAGEMENT IN 5G/6G NETWORKS

The network enablers including Software-Defined Networking (SDN), Network Function Virtualization (NFV), Mobile Edge Computing (MEC) can play an important role in the QoE-aware network management [14], [210]–[212]. The QoE management in 6G and beyond networks using network enabling technologies promises network programmability, scalability, agility, distributive computing, dynamic resource optimization and automation which will allow next-generation networks to fulfil the user-perceived quality for emerging video streaming applications while being cost-effective [12], [208], [213], [214]. The QoE management of the softwarized and virtualized next-generation networks requires the deployment of QoE monitoring and measurement solution on top of SDN controller/NFV Management and Orchestration (MANO) [215]–[218]. Moreover, the deployment of QoE-aware management approaches in 6G networks requires information exchange among the major players in multimedia streaming which includes Internet Service Providers (ISP) and Over-The-Top service providers (OTTs) [13], [208]. The ISP and OTT have different roles in multimedia service delivery and both have access and control of different information and resources. For example, an ISP has control over network infrastructure and access to network-level information based on which ISP can perform optimization of the network resources. While OTTs being service provider has the access to application-level KQIs and can optimize the video streaming delivery at client and server-side. To date, the multimedia services are encrypted by OTTs leading to no retrieval of OTT's application KQIs in the hand of ISP [219]–[221]. The exchange of application-level KQIs among OTTs and ISPs stands important for QoE management of the video streaming services as QoE is a multidisciplinary concept and most of the QoE models are mainly composed of the application KQIs such as video quality layers, average bitrate, stalling events, duration of the stalling event. Being blind to these KQIs, the ISPs cannot effectively perform QoE-aware network management [11], [222], [223]. The exchange of information among OTTs and ISP will require Service Level Agreements (SLAs) or Experience Level Agreements (ELA) [5]. Also, ELAs based on QoE Model will be needed

for offering services to the end customers [203], [224]. Furthermore, the E2E QoE management in the future networks will require a collaborative network and service management by both ISP and OTT provider based on the agreed QoE model and integration of the QoE model in the whole network and service delivery/optimization [11], [12], [224]. In the literature, several state-of-the-art works have been proposed for the collaborative QoE-aware network and service management where different business-centric QoE management strategies such as joint-venture, QoE-fairness based on fairness metric defined in [225], Customer Lifetime Value (CLV) and Zero-rated QoE has been proposed [11], [203], [208], [222], [224]. However, more future research work in the direction of the E2E collaborative QoE management of multimedia services by OTT-ISP collaboration is needed. The ongoing OTT-ISP collaboration in the industry can be seen in industrial projects such as T-Mobile Binge On where ISP limit the data rate of the OTT application and provide zero-rated data of the collaborating OTTs to its customers [226]. The Google Global Cache [227] and Netflix [228] Open Connects are other examples from industrial OTT-ISP provider collaborative service management where Google and Netflix allow collaborating ISP to host OTT's surrogate servers at the network edge to reduce content retrieval latency and unnecessary traffic from the core network of ISP. However, these approaches are not QoE-aware. One can argue upon the Network Neutrality on the formulation of the collaborative E2E QoE management approach which stands important in some geographical locations such as Europe [229]. Since June 2018, in United States Network Neutrality has been repealed [230] as it hinders the basic concepts of application-aware network slicing in 5G [231]. Therefore, a better definition of Network Neutrality in terms of QoE is required in the geographical location where it has been enforced [208].

B. QoE MONITORING

In the literature, many state-of-the-art works propose different passive and active QoE monitoring solutions based on user-end probes for the collection of the application-level QoE KQI at the user-end [232]–[240]. However, few works consider a cross/multi-layer model approach towards the QoE monitoring [232], [233]. Therefore, future research works for the cross/multi-layer QoE monitoring solutions are needed. The collection of the QoE related KQIs from the customer premises/user-device/terminal also triggers concerns about user privacy and security which needs further research in the domain of the QoE monitoring [2], [219], [241]. There are few efforts in the state-of-the-art where QoE KQIs are extracted from the encrypted OTT video streaming session [241]–[246]. However, further research in this direction in line with user privacy and security for QoE monitoring is needed.

The deployment of the QoE-aware network management approaches using network enabling technologies requires the standardized secure interfaces for QoE monitoring and

TABLE 10. Summary of research challenges and future direction in QoE model integration and QoE monitoring/management.

Research Topic	Research Challenges and Future Research Direction
QoE Management in 5G/6G networks	The network enabling technologies such as SDN, NFV and MEC allows the dynamic QoE-aware network resource allocation in the next-generation networks for the multimedia services. The deployment of the QoE management solutions on top of SDN controller/NFV MANO will deliver cost-effective optimization and automation of the network infrastructure. Indeed, the QoE-aware network and service management of the multimedia application requires information exchange among OTT and ISP. Therefore, future research requires the development of standardized and secure interfaces among OTT and ISP for sharing QoE KPIs, collaborative QoE-aware service management approaches and integration of ELAs for service delivery/business model. Additionally, standardization efforts are required to include QoE management in softwarized and virtualized next-networks.
QoE Monitoring	In the literature, the state-of-art works propose different passive/active QoE monitoring solutions using the user-end probe where few studies consider multi-layer QoE monitoring solutions. As QoE is a multidisciplinary concept, future research towards multi-layer QoE monitoring solutions in line with user privacy and security for monitoring QoE KPIs is need. Furthermore, future research requires the extension of the existing interfaces/data representation for secure and ultra-compressed data monitoring and finding an optimal tradeoff between monitoring frequency and control plane traffic.
QoE Model Integration & QoE measurements	Most of the QoE models in the literature consider only application-level KQI while ignoring the fact that QoE is a multidisciplinary concept. Therefore, future research should consider the development of the cross-layer QoE models. Further research is required to propose the long term and more accurate QoE models which can be easily deployed in the network management framework enabled by SDN and NFV MANO in next-generation networks.

the exchange of information from the end-user equipment, among Virtual Network Functions (VNFs), among inter/intra slice SDN controllers, and among ISP/MNOs and OTT providers [208], [247], [248]. The extension of web socket-based interfaces and Remote Procedure Calls (RPC) protocols can leverage the development of new standardized interfaces for data monitoring and information exchange to deploy data-driven network management. Additionally, finding the optimal trade-off between the control plane traffic and QoE monitoring frequency for optimization in 6G and beyond remains an open challenge. Future research in the direction of ultra compressed data representation formats can decrease the volume of control data [234], [249]. The QoE monitoring frequency depends on the time interval of optimization, accuracy/data requirements of optimization algorithm(s), and time-period for the optimization [250]. For example, if the optimization is being performed at O-RAN near-real-time RIC then extremely low latency is required which needs secure fast information retrieval as compared to the optimization being performed at O-RAN non-real-time RIC [251], [252].

The virtualized and softwarized network infrastructure in 5G/6G networks allows flexible deployment of the QoE monitoring solution using SDN and NFV but security, reliability, scalability and placement of the QoE monitoring solutions remain open questions for the research community [219], [253]–[255]. Furthermore, the cost-effectiveness such OPERational Expenditures (OPEX) and CAPital EXpenditures (CAPEX) for deploying QoE monitoring solutions in the next-generation 5G/6G networks using softwarized/virtualized probes remains an open challenge and question [12], [211], [256]–[258].

C. QoE MODEL INTEGRATION & QoE MEASUREMENTS

The QoE model integration is one of the most crucial parts for QoE-aware multimedia service delivery in the future Internet as QoE monitoring, measurement, management and optimization completely depends on it. For example, the QoE KQIs to be monitored by QoE monitoring solutions highly depends on the parameters of the QoE model that is being used. Mostly, QoE models in the literature consider

only application-level KQIs while ignoring the fact that QoE is a cross-disciplinary field that also includes multiple influencing factors such as network, business and system (end-user) devices [259], [260]. Therefore, future research is required for developing a cross-layer QoE model to cover most of the influencing factors on the user-perceived quality. Moreover, the video streaming QoE model found in the state-of-the-art works are mainly developed for the short time scale leading to less accurate QoE prediction for the long video streaming sessions. Hence, long/multi-time scale QoE model development for long video streaming session requires future research [261]–[263]. The deployment of the QoE models on top of SDN controller/NFV MANO is another open issue when it comes to network resource management of 5G/6G networks as monitoring QoE to effectively predicts may require additional interfaces in the future network architecture and it may create unnecessary overhead of control traffic in the control plane [208], [234], [264]. Future research should consider the time-varying nature of the QoE and time scale of different network resource optimization for developing new QoE models to be integrated into next-generation networks for QoE monitoring, measurement and management purposes leading to E2E QoE-aware multimedia service delivery.

D. SUMMARY

This section discusses the research challenges and future research directions for QoE-aware network and service management of multimedia streaming in the 5G/6G networks. Section VII-A highlights the major research challenges and opportunities related to QoE management of the video streaming application in 5G/6G networks including QoE-aware network management using SDN/NFV, deployment options for QoE-aware network optimization, collaborative network and service management by ISP and OTT, ELAs and QoE KQIs information exchange among OTT and ISP, ELAs based service agreements, and network neutrality and beyond. Section VII-B investigates the QoE monitoring challenges and future research directions in the next-generation networks such as multi-layer user-end probe-based QoE monitoring solutions, QoE monitoring

from encrypted OTT services and user privacy, secure interfaces for KQIs information retrieval, optimization/trade-off between control traffic and monitoring frequency and deployment of QoE monitoring solution in the virtualized and softwarized 5G/6G networks considering CAPEX/OPEX. Section VII-C discusses the QoE model integration challenges in the future networks including the development of the cross/multi-layer and long/multi time scale model considering different optimization intervals and different layers of network/radio protocol stack for network resource management in the 5G/6G networks. Table 10 provides the summary of the research challenges and future research direction of this section.

VIII. CONCLUSION

With the advancement of technologies such as adaptive HTTP streaming, the time dynamics of user QoE has become a crucial issue. In this paper, we provided a survey on video quality models and state-of-the-art works for global video quality measurement and CTVQ models. In addition, we also highlight ten research challenges and future directions that impact QoE modelling, monitoring, optimization, and management in the end-to-end multimedia service delivery chain involving next generation networks. These research challenges are grouped under three categories-1) Continuous Time-Varying QoE modelling, 2) QoE-aware video encoding and decoder optimization, and 3) QoE-aware network monitoring and service management.

Unlike predicting the overall quality of an image sequence, measuring time-varying subjective quality is a significant challenge. However, such a quality model is a must to achieve maximum QoE for end-users with emerging adaptive HTTP streaming applications. To create a predictive quality model considering the time-varying nature of the user-perceived quality, this paper first provides a discussion on state-of-the-art works considering spatial/temporal distortions, memory effect of HVS, effects of hysteresis for QoE, and use of pooling methods to combine spatial-temporal distortion measures. Moreover, a generic quality model is also presented to account for main factors affecting the user QoE in time-varying application scenarios followed by the research challenges for time-varying quality model development.

Next we discuss research challenges and future directions in QoE-aware multimedia creation, encoding and decoder optimization. We identify image/video compression, encoder/decoder optimization, error resilience, and concealment-aware video encoding as major factors that affect end user QoE in multimedia applications. With the emergence of novel video coding standards and multimedia technologies, their compression capabilities, complexity, and energy consumption have a significant impact on the end-user QoE. Hence, the adoption of machine learning technologies towards QoE-aware video coding, green multimedia technologies that reduce overall energy consumption in the multimedia delivery chain are discussed. Furthermore, we also emphasize the importance of QoE modelling and

its integration in the application layer at the content creation stage. This facilitates compressed video signals to become resilient to severe quality degradation due to transmission errors.

Finally, state-of-the-art works on QoE management, QoE monitoring and measurement, and model integration in 5G/6G networks are discussed followed by their respective research challenges and future directions. We emphasize the importance of QoE-aware network optimization, collaborative network, and service management in 5G/6G networks. We also discuss probe-based QoE monitoring, and cross layer model integration challenges in next generation networks.

Thus, this paper presents a comprehensive review of state-of-the-art works on time-varying user-perceived quality (QoE) modelling for video streaming applications and covers the full spectrum of QoE modelling across all stages of multimedia service delivery chain from the content creation, encoding, decoding, network and service management where the future research challenges and directions are discussed.

REFERENCES

- [1] Cisco. (Mar. 2020). *Cisco Visual Networking Index: Forecast and Methodology, 2018–2023*. Accessed: Mar. 28, 2022. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- [2] S. Baraković and L. Skarin-Kapov, "Survey and challenges of QoE management issues in wireless networks," *J. Comput. Netw. Commun.*, vol. 2013, pp. 1–28, Dec. 2013.
- [3] T. Hobfeld, R. Schatz, M. Varela, and C. Timmerer, "Challenges of QoE management for cloud applications," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 28–36, Apr. 2012.
- [4] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Höbfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 1st Quart., 2014.
- [5] M. Varela, P. Zwickl, P. Reichl, M. Xie, and H. Schulzrinne, "From service level agreements (SLA) to experience level agreements (ELA): The challenges of selling QoE to the user," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1741–1746.
- [6] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of HTTP adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2141–2153, Aug. 2016.
- [7] Q. Huynh-Thu, M.-N. Garcia, F. Speranza, P. Corriveau, and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 1–14, Mar. 2011.
- [8] M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnström, and A. Raake, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. 6th Int. Workshop Quality Multimedia Exp. (QoMEX)*, Sep. 2014, pp. 141–146.
- [9] R. R. R. Rao, S. Göring, and A. Raake, "Towards high resolution video quality assessment in the crowd," in *Proc. 13th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Jun. 2021, pp. 1–6.
- [10] R. Schatz, S. Schwarzmann, T. Zinner, O. Dobrijevic, E. Liotou, P. Pocta, S. Barakovic, J. B. Husic, and L. Skarin-Kapov, "QoE Management for future networks," in *Autonomous Control for a Reliable Internet of Services*. Cham, Switzerland: Springer, 2018, pp. 49–80.
- [11] A. Floris, A. Ahmad, and L. Atzori, "QoE-aware OTT-ISP collaboration in service management: Architecture and approaches," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 2s, pp. 1–24, Apr. 2018.
- [12] A. A. Barakabitze, N. Barman, A. Ahmad, S. Zadtootaghaj, L. Sun, M. G. Martini, and L. Atzori, "QoE management of multimedia streaming services in future networks: A tutorial and survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 526–565, 1st Quart., 2020.

- [13] A. Ahmad, A. Floris, and L. Atzori, "QoE-aware service delivery: A joint-venture approach for content and network providers," in *Proc. 8th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Jun. 2016, pp. 1–6.
- [14] A. Ahmad, A. Floris, and L. Atzori, "Timber: An SDN-based emulation platform for experimental research on video streaming," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1374–1387, Jul. 2020.
- [15] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [16] J. M. Boyce, Y. Yan, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34, Jan. 2016.
- [17] J. Nightingale, Q. Wang, and C. Grecos, "Scalable HEVC (SHVC)-based video stream adaptation in wireless networks," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 3573–3577.
- [18] S. Rakngan, T. Mallikarachi, and A. Fernando, "An adaptive video streaming framework for scalable HEVC (SHVC) standard," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–2.
- [19] *Preview Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats*, Standard ISO/IEC 23009-1:2019, 2019. Accessed: Dec. 16, 2021. [Online]. Available: <https://www.iso.org/standard/79329.html>
- [20] Apple. (2017). *Apple HTTP Live Streaming*. Accessed: Dec. 16, 2021. [Online]. Available: <https://developer.apple.com/streaming/>
- [21] Adobe. (2017). *Adobe HTTP Dynamic Streaming (HDS)*. Accessed: Dec. 16, 2021. [Online]. Available: <https://www.adobe.com/devnet/hds.html>
- [22] Microsoft. (2017). *Microsoft Silverlight Smooth Streaming*. Accessed: Dec. 16, 2021. [Online]. Available: <https://www.microsoft.com/silverlight/smoothstreaming/>
- [23] U. S. M. Dayananda and V. Swaminathan, "Investigating scalable high efficiency video coding for HTTP streaming," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun. 2015, pp. 1–6.
- [24] R. Deng and G. Liu, "QoE driven cross-layer scheme for DASH-based scalable video transmission over LTE," *Multimedia Tools Appl.*, vol. 77, no. 6, pp. 6445–6469, Mar. 2018.
- [25] C. Kreuzberger, D. Posch, and H. Hellwagner, "A scalable video coding dataset and toolchain for dynamic adaptive streaming over HTTP," in *Proc. 6th ACM Multimedia Syst. Conf. (MMSys)*. New York, NY, USA: Association for Computing Machinery, Mar. 2015, pp. 213–218.
- [26] H. Ebbinghaus, "Memory: A contribution to experimental psychology," *Ann. Neurosci.*, vol. 20, no. 4, p. 155, 2013.
- [27] I. Neath and A. M. Surprenant, *Human Memory: An Introduction to Research, Data, and Theory*. Belmont, CA, USA: Wadsworth, 2003.
- [28] D. E. Pearson, "Viewer response to time-varying video quality," *Proc. SPIE*, vol. 3299, pp. 16–25, Jul. 1998.
- [29] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–292, 1979.
- [30] K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Process.*, vol. 70, no. 3, pp. 279–294, 1998.
- [31] S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. García, "Quality of experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology," *Signal Process., Image Commun.*, vol. 39, pp. 432–443, Nov. 2015.
- [32] K. Seshadrinathan and A. C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Prague, Czech Republic, May 2011, pp. 1153–1156.
- [33] Z. He, S. Mao, and T. Jiang, "A survey of QoE-driven video streaming over cognitive radio networks," *IEEE Netw.*, vol. 29, no. 6, pp. 20–25, Nov/Dec. 2015.
- [34] M. T. Vega, C. Perra, F. De Turck, and A. Liotta, "A review of predictive quality of experience management in video streaming services," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 432–445, Jun. 2018.
- [35] A. Ahmad, A. B. Mansoor, A. A. Barakabitze, A. Hines, L. Atzori, and R. Walshe, "Supervised-learning-based QoE prediction of video streaming in future networks: A tutorial with comparative study," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 88–94, Nov. 2021.
- [36] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 285–302, 1st Quart., 2017.
- [37] N. Barman and M. G. Martini, "QoE modeling for HTTP adaptive video streaming—A survey and open challenges," *IEEE Access*, vol. 7, pp. 30831–30859, 2019.
- [38] T. Hoßfeld, R. Schatz, S. Biedermann, A. Platzer, S. Egger, and M. Fiedler, "The memory effect and its implications on web QoE modeling," in *Proc. 23rd Int. Teletraffic Congr. (ITC)*, San Francisco, CA, USA, Sep. 2011, pp. 103–110.
- [39] D. S. Hands and S. E. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Appl. Cogn. Psychol.*, vol. 15, no. 6, pp. 639–657, Nov. 2001.
- [40] A. S. Luchins, "Primacy-recency in impression formation," in *The Order of Presentation in Persuasion*, C. I. Hovland, Ed. New Haven, CT, USA: Yale Univ. Press, 1957, pp. 33–61.
- [41] L. Gros and N. Chateau, "Instantaneous and overall judgements for time-varying speech quality: Assessments and relationships," *Acta Acustica United Acustica*, vol. 87, no. 3, pp. 367–377, 2001.
- [42] S. Voran, "A basic experiment on time-varying speech quality," in *Proc. 4th Int. Meas. Speech Audio Quality Netw. Conf. (MESAQIN)*, Prague, Czech Republic, Jun. 2005, pp. 1–14.
- [43] A. Raake, "Short- and long-term packet loss behavior: Towards speech quality prediction for arbitrary loss distributions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 1957–1968, Nov. 2006.
- [44] Z. Chen and H. Nakazato, "Model for time-varying quality of speech services," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, St. Louis, MO, USA, Nov. 2005, p. 5.
- [45] *Method for Objective Measurements of Perceived Audio Quality*, document ITU-R BS.1387, 2001.
- [46] B. Weiss, S. Möller, A. Raake, J. Berger, and R. Ullmann, "Modeling call quality for time-varying transmission characteristics using simulated conversational structures," *Acta Acustica United Acustica*, vol. 95, no. 6, pp. 1140–1151, Nov. 2009.
- [47] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage.*, May 2011, pp. 485–492.
- [48] S. Tavakoli, J. Gutiérrez, and N. García, "Subjective quality study of adaptive streaming of monoscopic and stereoscopic video," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 684–692, Apr. 2014.
- [49] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, Jr., and A. C. Bovik, "Modeling the time-varying subjective quality of HTTP video streams with rate adaptations," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2206–2221, May 2014.
- [50] K. Seshadrinathan and A. Bovik, "Video quality assessment," in *The Essential Guide to Video Processing*, no. 6, 2nd ed. Boston, MA, USA: Academic, 2009, ch. 14, pp. 417–436–1968.
- [51] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "No-reference quality assessment of tone-mapped HDR pictures," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2957–2971, Jun. 2017.
- [52] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1333–1337, Sep. 2017.
- [53] Netflix. (Jun. 2016). *Toward a Practical Perceptual Video Quality Metric*. Accessed: Dec. 16, 2021. [Online]. Available: <https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652>
- [54] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [55] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.
- [56] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [57] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 40:1–40:14, Jul. 2011.
- [58] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
- [59] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

- [60] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
- [61] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [62] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*. New York, NY, USA: ACM, 1993.
- [63] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, USA, Oct. 2006, pp. 2945–2948.
- [64] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "Video quality pooling adaptive to perceptual distortion severity," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 610–620, Feb. 2013.
- [65] J. You, A. Perkiš, M. M. Hannuksela, and M. Gabbouj, "Perceptual quality assessment based on visual attention analysis," in *Proc. 17th ACM Int. Conf. Multimedia*, Beijing, China, 2009, pp. 561–564.
- [66] S. Rimac-Drlje, M. Vranješ, and D. Žagar, "Influence of temporal pooling method on the objective video quality evaluation," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, May 2009, pp. 1–5.
- [67] Z. Tu, C.-J. Chen, L.-H. Chen, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "A comparative evaluation of temporal pooling methods for blind video quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 141–145.
- [68] Z. Duanmu, K. Ma, and Z. Wang, "Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 6135–6146, Dec. 2018.
- [69] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "Streaming video QoE modeling and prediction: A long short-term memory approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 661–673, Mar. 2020.
- [70] H. T. Tran, N. P. Ngoc, T. Hoßfeld, M. Seufert, and T. C. Thang, "Cumulative quality modeling for HTTP adaptive streaming," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 1, pp. 1–24, 2021.
- [71] *LIVE Video Quality Database*, Univ. Texas Austin, Austin, TX, USA, 2009. Accessed: Dec. 16, 2021.
- [72] EPFL-PoliMI. (2009). *EPFL-PoliMI Video Quality Assessment Database*. Accessed: Dec. 16, 2021. [Online]. Available: <http://vqa.com.polimi.it/>
- [73] M. Vranješ, S. Rimac-Drlje, and D. Žagar, "Subjective and objective quality evaluation of the H.264/AVC coded video," in *Proc. 15th Int. Conf. Syst., Signals Image Process.*, Bratislava, Slovakia, Jun. 2008, pp. 287–290.
- [74] S. Rimac-Drlje, M. Vranješ, and D. Žagar, "Influence of temporal pooling method on the objective video quality evaluation," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, May 2009, pp. 1–5.
- [75] J. You, J. Korhonen, and A. Perkiš, "Attention modeling for video quality assessment: Balancing global quality and local quality," in *Proc. IEEE Int. Conf. Multimedia Expo*, Singapore: Suntec City, Jul. 2010, pp. 914–919.
- [76] K. Lee, J. Park, S. Lee, and A. C. Bovik, "Temporal pooling of video quality estimates using perceptual motion models," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 2493–2496.
- [77] J. You, J. Korhonen, and A. Perkiš, "Spatial and temporal pooling of image quality metrics for perceptual video quality assessment on packet loss streams," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 1002–1005.
- [78] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.
- [79] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proc. 3rd Int. Workshop Video Process. Quality Metrics (CD-ROM)*, 2007, p. 4.
- [80] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To pool or not to pool: A comparison of temporal pooling methods for HTTP adaptive video streaming," in *Proc. 15th Int. Workshop Quality Multimedia Exper. (QoMEX)*, Jul. 2013, pp. 52–57.
- [81] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. Möller, "An objective and subjective quality assessment study of passive gaming video streaming," *Int. J. Netw. Manage.*, vol. 30, no. 3, p. 2054, May 2020.
- [82] Netflix. (Oct. 2018). *VMAF: The Journey Continues*. Accessed: Dec. 16, 2021. [Online]. Available: <https://medium.com/netflix-techblog/vmaf-the-journey-continues-44b51ee9ed12>
- [83] M. Barkowsky, B. Eskofier, R. Bitto, J. Bialkowski, and A. Kaup, "Perceptually motivated spatial and temporal integration of pixel based video quality measures," in *Proc. Welcome to Mobile Content Quality Exp. (MobConQoE)*, New York, NY, USA, 2007, pp. 4:1–4:7.
- [84] C. J. van den Lambrecht, "Perceptual models and architectures for video coding applications," M. S. thesis, EPFL, Lausanne, Switzerland, 1996.
- [85] A. B. Watson, Q. J. Hu, I. J. F. McGowan, and J. B. Mulligan, "Design and performance of a digital video quality metric," *Proc. SPIE*, vol. 3644, pp. 168–174, May 1999.
- [86] J. Lubin, M. Brill, A. de Vries, and O. Finard, "Method and apparatus for assessing the visibility of differences between two image sequences," U.S. Patent 5 974 159, Oct. 26, 1999.
- [87] S. Winkler, "Perceptual distortion metric for digital color video," *Proc. SPIE*, vol. 3644, pp. 175–184, May 1999.
- [88] D. Ghadiyaram, J. Pan, and A. C. Bovik, "Learning a continuous-time streaming video QoE model," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2257–2271, May 2018.
- [89] T. Nunome and K. Abe, "Relationships between usage situations and spatiotemporal QoE in audio and video IP transmission," in *Proc. 3rd Int. Conf. Comput. Commun. Syst. (ICCCS)*, Apr. 2018, pp. 278–281.
- [90] M. Masry and S. S. Hemami, "Perceived quality metrics for low bit rate compressed video," in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, USA, Sep. 2002, p. III.
- [91] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1316–1327, Sep. 2006.
- [92] T. Stockhammer, "Dynamic adaptive streaming over HTTP: Standards and design principles," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, Santa Clara, CA, USA, Feb. 2011, pp. 133–144.
- [93] M. A. Masry and S. S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 133–146, Feb. 2004.
- [94] N. Eswara, D. S. V. Reddy, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "A linear regression framework for assessing time-varying subjective quality in HTTP streaming," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Montreal, QC, Canada, Nov. 2017, pp. 31–35.
- [95] C. T. E. R. Hewage and M. G. Martini, "Time varying quality estimation for HTTP based adaptive video streaming," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6.
- [96] H. T. T. Tran, N. P. Ngoc, T. Hoßfeld, M. Seufert, and T. C. Thang, "Cumulative quality modeling for HTTP adaptive streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, pp. 1–24, Apr. 2021.
- [97] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 154–166, Feb. 2017.
- [98] N. Eswara, K. Manasa, A. Kommineni, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "A continuous QoE evaluation framework for video streaming over HTTP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3236–3250, Nov. 2018.
- [99] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5217–5231, Nov. 2017.
- [100] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A subjective and objective study of stalling events in mobile streaming videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 183–197, Jan. 2019.
- [101] C. Bampis, Z. Li, I. Katsavounidis, T.-Y. Huang, C. Ekanadham, and A. Bovik, "Towards perceptually optimized end-to-end adaptive video streaming," 2018, [arXiv:1808.03898](https://arxiv.org/abs/1808.03898).
- [102] Y. Kawayoke and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 385–388.
- [103] C. Chen, X. Zhu, G. de Veciana, A. Bovik, and R. W. Heath, Jr., "Adaptive video transmission with subjective quality constraints," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2477–2481.

- [104] T. Kim, K. Lee, S. Lee, and A. C. Bovik, "A new multimodal interactive way of subjective scoring of 3D video quality of experience," *Proc. SPIE*, vol. 9011, Mar. 2014, Art. no. 901127.
- [105] European Commission-FP7. (Feb. 2015). *Content and Context Aware Delivery for Interactive Multimedia Healthcare Applications (CONCERTO)*. Accessed: Dec. 16, 2021. [Online]. Available: <https://cordis.europa.eu/project/id/288502/>
- [106] C. G. Bampis, Z. Li, I. Katsavounidis, T.-Y. Huang, C. Ekanadham, and A. C. Bovik, "Psychopy-based subjective interface demo," Dept. Elect. Comput. Eng., Lab. Image Video Eng., Univ. Texas Austin, Austin, TX, USA, Tech. Rep. Accessed: Dec. 16, 2021.
- [107] *Dynamic Adaptive Streaming Over HTTP (DASH)—Part 5: Server and Network Assisted DASH (SAND)*, ISO/IEC 23009-5:2017, 2017.
- [108] *Methodologies for the Subjective Assessment of the Quality of Television Images*, document ITU-R BT.500-14, Oct. 2019.
- [109] *Parametric Bitstream-Based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services Over Reliable Transport*, ITU, Geneva, Switzerland, 2017. Accessed: Dec. 16, 2021.
- [110] *Video Quality Assessment of Streaming Services Over Reliable Transport for Resolutions up to 4 K*, ITU, Geneva, Switzerland, 2020. Accessed: Dec. 16, 2021.
- [111] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1," in *Proc. 9th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Erfurt, Germany, May 2017, p. 1203.
- [112] W. Robitza, S. Göring, A. Raake, D. Lindgren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M.-N. Garcia, K. Yamagishi, and S. Broom, "HTTP adaptive streaming QoE estimation with ITU-T rec. P. 1203: Open databases and software," in *Proc. 9th ACM Multimedia Syst. Conf.*, Amsterdam, The Netherlands, Jun. 2018, p. 1203.
- [113] R. Perera, A. Fernando, T. Mallikarachchi, H. K. Arachchi, and M. Pourazad, "QoE aware resource allocation for video communications over LTE based mobile networks," in *Proc. 10th Int. Conf. Heterogeneous Netw. Quality, Rel., Secur. Robustness*, Aug. 2014, pp. 63–69.
- [114] M. Gadaleta, F. Chiariotti, M. Rossi, and A. Zanella, "D-dash: A deep Q-learning framework for DASH video streaming," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 703–718, Dec. 2017.
- [115] D. J. Vergados, A. Michalakis, A. Sgora, D. D. Vergados, and P. Chatzimisios, "FDASH: A fuzzy-based MPEG/DASH adaptation algorithm," *IEEE Syst. J.*, vol. 10, no. 2, pp. 859–868, Jun. 2016.
- [116] C. Zhou, C.-W. Lin, and Z. Guo, "mDASH: A Markov decision-based rate adaptation approach for dynamic HTTP streaming," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 738–751, Apr. 2016.
- [117] L. Zhou, J. J. P. C. Rodrigues, H. Wang, M. Martini, and V. C. M. Leung, "5G multimedia communications: Theory, technology, and application," *IEEE Multimedia Mag.*, vol. 26, no. 1, pp. 8–9, Jan. 2019.
- [118] Cisco. (Jun. 2017). *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*. Accessed: Dec. 16, 2021. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf>
- [119] T. Ebrahimi and M. Kunt, "Visual data compression for multimedia applications," *Proc. IEEE*, vol. 86, no. 6, pp. 1109–1125, Jun. 1998.
- [120] M. Ghanbari, *Video Coding: An Introduction to Standard Codecs*. London, U.K.: Institution of Electrical Engineers, 1999.
- [121] S. Vivienne, B. Madhukar, and S. J. Gary, *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, 2014th ed. Cham, Switzerland: Springer, Aug. 2014.
- [122] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [123] S. Andrew, B. Vittorio, B. Jill, C. Jianle, and S. Teruhiko, "Joint call for proposals on video compression with capability beyond HEVC," Joint Video Explor. Team ITU-T SG 16 WP 3, ISO/IEC JTC 1/SC 29/WG 11, Tech. Rep. JVET-H 002 (v5), 2017.
- [124] B. Bross, J. Chen, S. Liu, and Y. Wang, *Versatile Video Coding (Draft 10)*, document JVET-S2001, Joint Video Experts Team (JVET), 2020.
- [125] D. Mukherjee, J. Han, J. Bankoski, R. Bultje, A. Grange, J. Koleszar, P. Wilkins, and Y. Xu, "A technical overview of VP9—The latest open-source video codec," in *Proc. SMPTE Annu. Tech. Conf. Exhib.*, Oct. 2013, pp. 1–17.
- [126] Y. Chen *et al.*, "An overview of core coding tools in the AV1 video codec," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2018, pp. 41–45.
- [127] I. E. Richardson, *The H.264 Advanced Video Compression Standard*, 2nd ed. Hoboken, NJ, USA: Wiley, Apr. 2010.
- [128] L. Lin, S. Yu, T. Zhao, and Z. Wang, "PEA265: Perceptual assessment of video compression artifacts," *CoRR*, vol. abs/1903.00473, pp. 1–11, Mar. 2019.
- [129] B. Li, H. Li, L. Li, and J. Zhang, "λ domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sep. 2014.
- [130] *HEVC Test Model Version—16.16*, JCT-VC, Macao, 2017. Accessed: Dec. 16, 2021.
- [131] Z. Liu, X. Pan, Y. Li, and Z. Chen, "A game theory based CTU-level bit allocation scheme for HEVC region of interest coding," *IEEE Trans. Image Process.*, vol. 30, pp. 794–805, 2021.
- [132] W. Gao, S. Kwong, and Y. Jia, "Joint machine learning and game theory for rate control in high efficiency video coding," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6074–6089, Dec. 2017.
- [133] W. Gao, S. Kwong, Y. Zhou, and H. Yuan, "SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 988–999, Jun. 2016.
- [134] M. H. Hyun, B. Lee, and M. Kim, "A frame-level constant bit-rate control using recursive Bayesian estimation for versatile video coding," *IEEE Access*, vol. 8, pp. 227255–227269, 2020.
- [135] S. Zhu, C. Liu, and Z. Xu, "High-definition video compression system based on perception guidance of salient information of a convolutional neural network and HEVC compression domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 1946–1959, Jul. 2019.
- [136] X. Sun, H. Ma, W. Zuo, and M. Liu, "Perceptual-based HEVC intra coding optimization using deep convolution networks," *IEEE Access*, vol. 7, pp. 56308–56316, 2019.
- [137] Z. Zhang, T. Jing, J. Han, Y. Xu, and F. Zhang, "A new rate control scheme for video coding based on region of interest," *IEEE Access*, vol. 5, pp. 13677–13688, 2017.
- [138] J. Ascenso, P. Akyazi, F. Pereira, and T. Ebrahimi, "Learning-based image coding: Early solutions reviewing and subjective quality evaluation," *Proc. SPIE*, vol. 11353, pp. 164–176, Apr. 2020.
- [139] Y. Y. J. Chen and S. H. Kim, "Algorithm description for versatile video coding and test model 8 (VTM 8)," Joint Video Experts Team (JVET), Tech. Rep. JVET-R_Notes_d1, 2020.
- [140] J. Han, B. Li, D. Mukherjee, C. H. Chiang, A. Grange, C. Chen, H. Su, S. Parker, S. Deng, U. Joshi, Y. Chen, Y. Wang, P. Wilkins, Y. Xu, and J. Bankoski, "A technical overview of AV1," *Proc. IEEE*, vol. 109, no. 9, pp. 1435–1462, Sep. 2021.
- [141] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.
- [142] H. O. Dominguez and K. R. Rao, *Versatile Video Coding Latest Advances in Video Coding Standards*. Gistrup, Denmark: River Publishers, 2018.
- [143] P. Topiwala, M. Krishnan, and W. Dai, "Performance comparison of VVC, AV1, and HEVC on 8-bit and 10-bit content," *Proc. SPIE*, vol. 10752, pp. 305–314, Sep. 2018.
- [144] D. García-Lucas, G. Cebrián-Márquez, A. J. Díaz-Honrubia, T. Mallikarachchi, and P. Cuenca, "Cost-efficient HEVC-based quadtree splitting (HEQUS) for VVC video transcoding," *Signal Process., Image Commun.*, vol. 94, May 2021, Art. no. 116199.
- [145] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1668–1682, Jun. 2020.
- [146] Q. Zhang, Y. Zhao, B. Jiang, L. Huang, and T. Wei, "Fast CU partition decision method based on texture characteristics for H.266/VVC," *IEEE Access*, vol. 8, pp. 203516–203524, 2020.
- [147] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
- [148] B. Erabadda, T. Mallikarachchi, G. Kulupana, and A. Fernando, "ICUS: Intelligent CU size selection for HEVC inter prediction," *IEEE Access*, vol. 8, pp. 141143–141158, 2020.
- [149] T. Mallikarachchi, A. Fernando, and H. K. Arachchi, "Efficient coding unit size selection based on texture analysis for HEVC intra prediction," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
- [150] T. Mallikarachchi, D. S. Talagala, H. K. Arachchi, and A. Fernando, "Content-adaptive feature-based CU size prediction for fast low-delay video encoding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 693–705, Mar. 2018.

- [151] S.-H. Park and J.-W. Kang, "Fast affine motion estimation for versatile video coding (VVC) encoding," *IEEE Access*, vol. 7, pp. 158075–158084, 2019.
- [152] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 465–470, Feb. 2013.
- [153] F. Pakdaman, L. Yu, M. R. Hashemi, M. Ghanbari, and M. Gabbouj, "SVM based approach for complexity control of HEVC intra coding," *Signal Process., Image Commun.*, vol. 93, Apr. 2021, Art. no. 116177.
- [154] S. Ryu and J. Kang, "Machine learning-based fast angular prediction mode decision technique in video coding," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5525–5538, Nov. 2018.
- [155] G. Correa, P. A. Assuncao, L. V. Agostini, and L. A. da Silva Cruz, "Fast HEVC encoding decisions using data mining," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 660–673, Apr. 2015.
- [156] J. Xu, M. Xu, Y. Wei, Z. Wang, and Z. Guan, "Fast H.264 to HEVC transcoding: A deep learning method," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1633–1645, Jul. 2019.
- [157] F. Zaki, A. E. Mohamed, and S. G. Sayed, "CuNet: A deep learning-based framework for fast CTU partitioning of H265/HEVC intra-coding," *Ain Shams Eng. J.*, vol. 12, no. 2, pp. 1859–1866, Jun. 2021.
- [158] JVET. *VVC Test Model Version—2.0.1*. Accessed: Dec. 16, 2021. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-2.0.1
- [159] T. Mallikarachi, D. S. Talagala, H. K. Arachchi, and A. Fernando, "Decoding-complexity-aware HEVC encoding using a complexity–rate–distortion model," *IEEE Trans. Consum. Electron.*, vol. 64, no. 1, pp. 35–43, Feb. 2018.
- [160] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1685–1696, Dec. 2012.
- [161] L. Zou, R. Trestian, and G.-M. Muntean, "E3DOAS: Balancing QoE and energy-saving for multi-device adaptation in future mobile wireless video delivery," *IEEE Trans. Broadcast.*, vol. 64, no. 1, pp. 26–40, Mar. 2018.
- [162] A. M. Groba, P. J. Lobo, and M. Chavarrias, "QoE-aware dual control system to guarantee battery lifetime for mobile video applications," *IEEE Trans. Consum. Electron.*, vol. 65, no. 4, pp. 454–463, Nov. 2019.
- [163] M. A. Hoque, M. Siekinen, and J. K. Nurminen, "Energy efficient multimedia streaming to mobile devices—A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 579–597, 1st Quart., 2014.
- [164] T. Ma, M. Hempel, D. Peng, and H. Sharif, "A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 963–972, 3rd Quart., 2013.
- [165] C. Schurgers, V. Raghunathan, and M. B. Srivastava, "Power management for energy-aware communication systems," *ACM Trans. Embedded Comput. Syst.*, vol. 2, no. 3, pp. 431–447, Aug. 2003.
- [166] X. Pérez-Costa and D. Camps-Mur, "IEEE 802.11E QoS and power saving features overview and analysis of combined performance," *IEEE Wireless Commun.*, vol. 17, no. 4, pp. 88–96, Aug. 2010.
- [167] R. Wang, J. Tsai, C. Maciocco, T.-Y. C. Tai, and J. Wu, "Reducing power consumption for mobile platforms via adaptive traffic coalescing," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1618–1629, Sep. 2011.
- [168] R. Trestian, O. Ormond, and G.-M. Muntean, "Energy–quality–cost tradeoff in a multimedia-based heterogeneous wireless network environment," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 340–357, Jun. 2013.
- [169] E. Noguees, S. Holmbacka, M. Pelcat, D. Menard, and J. Lilius, "Power-aware HEVC decoding with tunable image quality," in *Proc. IEEE Workshop Signal Process. Syst. (SIPS)*, Oct. 2014, pp. 1–6.
- [170] L. Zou, R. Trestian, and G.-M. Muntean, "eDOAS: Energy-aware device-oriented adaptive multimedia scheme for Wi-Fi offload," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 2916–2921.
- [171] Q. Huynh-Thu and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 641–651, Sep. 2008.
- [172] E. Kim, H. Jeong, J. Yang, and M. Song, "Balancing energy use against video quality in mobile devices," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 517–524, Aug. 2014.
- [173] Y.-L. Chen, M.-F. Chang, and W.-Y. Liang, "Energy-efficient video decoding schemes for embedded handheld devices," *Multimedia Tools Appl.*, vol. 75, no. 6, pp. 3281–3300, Mar. 2016.
- [174] F. C. Fernandes, X. Ducloux, Z. Ma, E. Faramarzi, P. Gendron, and J. Wen, "The green metadata standard for energy-efficient video consumption," *IEEE Trans. Multimedia*, vol. 22, no. 1, pp. 80–87, Jan./Mar. 2015.
- [175] *Call for Proposals on Green MPEG*, MPEG, Joint Collaborative Team Video Coding (JCT-VC), Incheon, South Korea, 2013.
- [176] E. Noguees, S. Holmbacka, M. Pelcat, D. Menard, and J. Lilius, "Power-aware HEVC decoding with tunable image quality," in *Proc. IEEE Workshop Signal Process. Syst. (SIPS)*, Belfast, U.K., Oct. 2014, pp. 1–6.
- [177] E. Noguees, R. Berrada, M. Pelcat, D. Menard, and E. Raffin, "A DVFS based HEVC decoder for energy-efficient software implementation on embedded processors," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Turin, Italy, Jun. 2015, pp. 1–6.
- [178] Y. Benmoussa, E. Senn, N. Derouineau, N. Tizon, and J. Boukhobza, "Green metadata based adaptive DVFS for energy efficient video decoding," in *Proc. 26th Int. Workshop Power Timing Modeling, Optim. Simulation (PATMOS)*, Bremen, Germany, Sep. 2016, pp. 235–242.
- [179] M. G. Martini, C. W. Chen, Z. Chen, T. Dagiuklas, L. Sun, and X. Zhu, "Guest editorial QoE-aware wireless multimedia systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1153–1156, Aug. 2012.
- [180] V. DeBrunner, L. DeBrunner, L. Wang, and S. Radhakrishnan, "Error control and concealment for image transmission," *IEEE Commun. Surveys Tuts.*, vol. 3, no. 1, pp. 2–9, 1st Quart., 2000.
- [181] S. K. Bandyopadhyay, Z. Wu, P. Pandit, and J. M. Boyce, "An error concealment scheme for entire frame losses for H.264/AVC," in *Proc. IEEE Sarnoff Symp.*, Mar. 2006, pp. 1–4.
- [182] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "Video adaptation for consumer devices: Opportunities and challenges offered by new standards," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 157–163, Dec. 2014.
- [183] H.-S. Kang, Y.-W. Kim, and T.-Y. Kim, "A temporal error concealment method based on edge adaptive boundary matching," in *Advances in Image and Video Technology*, L.-W. Chang and W.-N. Lie, Eds. Berlin, Germany: Springer, 2006, pp. 852–860.
- [184] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Apr. 1993, pp. 417–420.
- [185] S. M. Marvasti-Zadeh, H. Ghanei-Yakhdan, and S. Kasaei, "An efficient adaptive boundary matching algorithm for video error concealment," *Iranian J. Elect. Electron. Eng.*, vol. 10, no. 3, pp. 188–202, 2014.
- [186] C. Liu, R. Ma, and Z. Zhang, "Error concealment for whole frame loss in HEVC," in *Advances on Digital Television and Wireless Multimedia Communications*, W. Zhang, X. Yang, Z. Xu, P. An, Q. Liu, and Y. Lu, Eds. Berlin, Germany: Springer, 2012, pp. 271–277.
- [187] J. Nightingale, Q. Wang, and C. Grecos, "Priority-based methods for reducing the impact of packet loss on HEVC encoded video streams," *Proc. SPIE*, vol. 8656, Feb. 2013, Art. no. 86560.
- [188] A. Jassal, B. Oztas, M. T. Pourazad, and P. Nasiopoulos, "A packet prioritization scheme for 3D-HEVC content transmission over LTE networks," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1788–1793.
- [189] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
- [190] W.-W. Xu, R.-X. Jiang, X. Tian, and Y.-W. Chen, "Content based error resilience coding using redundant coding, reference frame modification, and mode decision," *J. Central South Univ.*, vol. 21, no. 11, pp. 4199–4209, Nov. 2014.
- [191] A. Jerbi, J. Wang, and S. Shirani, "Error-resilient region-of-interest video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1175–1181, Sep. 2005.
- [192] P. Sivanantharasa, W. A. C. Fernando, and H. K. Arachchi, "Region of interest video coding with flexible macroblock ordering," in *Proc. 1st Int. Conf. Ind. Inf. Syst.*, Aug. 2006, pp. 596–599.
- [193] H. Hadizadeh and I. V. Bajic, "Burst-Loss-Resilient packetization of video," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3195–3206, Nov. 2011.
- [194] B. Katz, S. Greenberg, N. Yarkoni, N. Blaunsten, and R. Giladi, "New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 308–319, Mar. 2007.

- [195] J. F. M. Carreira, P. A. Assunção, S. M. M. de Faria, E. Ekmekcioglu, and A. Kondo, "A two-stage approach for robust HEVC coding and streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1960–1973, Aug. 2018.
- [196] J. Carreira, E. Ekmekcioglu, A. Kondo, P. Assuncao, S. Faria, and V. De Silva, "Selective motion vector redundancies for improved error resilience in HEVC," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2457–2461.
- [197] G. Kulupana, D. S. Talagala, H. K. Arachchi, and A. Fernando, "End user video quality prediction and coding parameters selection at the encoder for robust HEVC video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 11, pp. 3367–3381, Nov. 2019.
- [198] J. F. M. Carreira, P. A. Assunção, S. M. M. de Faria, E. Ekmekcioglu, and A. Kondo, "Error concealment-aware encoding for robust video transmission," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 282–293, Jun. 2019.
- [199] G. Kulupana, D. S. Talagala, H. K. Arachchi, M. Akinola, and A. Fernando, "Concealment support and error resilience for HEVC to improve consumer quality of experience," *IEEE Trans. Consum. Electron.*, vol. 67, no. 2, pp. 107–118, May 2021.
- [200] T. Höbfeld, P. E. Heegaard, M. Varela, L. Skorin-Kapov, and M. Fiedler, "From QoS distributions to QoE distributions: A system's perspective," in *Proc. 6th IEEE Conf. Netw. Softwarization (NetSoft)*, Jun. 2020, pp. 51–56.
- [201] F. Loh, F. Poignée, F. Wamser, F. Leidinger, and T. Höbfeld, "Uplink vs. downlink: Machine learning-based quality prediction for HTTP adaptive video streaming," *Sensors*, vol. 21, no. 12, p. 4172, Jun. 2021.
- [202] J. Seppänen, M. Varela, and A. Sgora, "An autonomous QoE-driven network management framework," *J. Vis. Commun. Image Represent.*, vol. 25, no. 3, pp. 565–577, Apr. 2014.
- [203] A. Ahmad, "QoE-aware multimedia service management and monitoring through OTT and ISP collaboration," Ph.D. dissertation, Università degli Studi di Cagliari, Cagliari, Italy, 2019.
- [204] P. Casas and R. Schatz, "Quality of experience in cloud services: Survey and measurements," *Comput. Netw.*, vol. 68, pp. 149–165, Aug. 2014.
- [205] R. Schatz, T. Höbfeld, and P. Casas, "Passive YouTube QoE monitoring for ISPs," in *Proc. 6th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Jul. 2012, pp. 358–364.
- [206] H. Zarini, A. Khalili, H. Tabassum, and M. Rasti, "Joint transmission in QoE-driven backhaul-aware MC-NOMA cognitive radio network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.
- [207] I.-S. Comşa, A. Molnar, I. Tal, P. Bergamin, G.-M. Muntean, C. H. Muntean, and R. Treştian, "A machine learning resource allocation solution to improve video quality in remote education," *IEEE Trans. Broadcast.*, vol. 67, no. 3, pp. 664–684, Sep. 2021.
- [208] A. Ahmad and L. Atzori, "MNO-OTT collaborative video streaming in 5G: The zero-rated QoE approach for quality and resource management," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 1, pp. 361–374, Mar. 2020.
- [209] L. Skorin-Kapov and M. Varela, "A multi-dimensional view of QoE: The ARCU model," in *Proc. 35th Int. Conf. MIPRO*, 2012, pp. 662–666.
- [210] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 3rd Quart., 2018.
- [211] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges," *Comput. Netw.*, vol. 167, Feb. 2020, Art. no. 106984.
- [212] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.
- [213] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
- [214] M. Hoffmann, M. Jarschel, R. Pries, P. Schneider, A. Jukan, W. Bziuk, S. Gebert, T. Zinner, and P. Tran-Gia, "SDN and NFV as enabler for the distributed network cloud," *Mobile Netw. Appl.*, vol. 23, no. 3, pp. 521–528, 2018.
- [215] J. C. Fernandez, T. Taleb, M. Guizani, and N. Kato, "Bandwidth aggregation-aware dynamic QoS negotiation for real-time video streaming in next-generation wireless networks," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1082–1093, Oct. 2009.
- [216] A. Laghrissi and T. Taleb, "A survey on the placement of virtual resources and virtual network functions," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1409–1434, 2nd Quart., 2019.
- [217] I. Afolabi, M. Bagaa, W. Boumezer, and T. Taleb, "Toward a real deployment of network services orchestration and configuration convergence framework for 5G network slices," *IEEE Netw.*, vol. 35, no. 1, pp. 242–250, Jan. 2021.
- [218] A. N. Abbou, T. Taleb, and J. Song, "A software-defined queuing framework for QoS provisioning in 5G and beyond mobile systems," *IEEE Netw.*, vol. 35, no. 2, pp. 168–173, Mar. 2021.
- [219] W. Robitza, A. Ahmad, P. A. Kara, L. Atzori, M. G. Martini, A. Raake, and L. Sun, "Challenges of future multimedia QoE monitoring for internet service providers," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 22243–22266, Nov. 2017.
- [220] F. Wamser, D. Hock, M. Seufert, B. Staehle, R. Pries, and P. Tran-Gia, "Using buffered playtime for QoE-oriented resource management of YouTube video streaming," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 3, pp. 288–302, Apr. 2013.
- [221] M. Seufert, P. Casas, N. Wehner, L. Gang, and K. Li, "Stream-based machine learning for real-time QoE analysis of encrypted video streaming traffic," in *Proc. 22nd Conf. Innov. Clouds, Internet Netw. Workshops (ICIN)*, Feb. 2019, pp. 76–81.
- [222] A. Ahmad, A. Floris, and L. Atzori, "QoE-aware service delivery: A joint-venture approach for content and network providers," in *Proc. 8th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Jun. 2016, pp. 1–6.
- [223] L. Skorin-Kapov, M. Varela, T. Höbfeld, and K.-T. Chen, "A survey of emerging concepts and challenges for QoE management of multimedia services," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 2s, pp. 1–29, Apr. 2018.
- [224] A. Ahmad, A. Floris, and L. Atzori, "OTT-ISP joint service management: A customer lifetime value based approach," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manage. (IM)*, May 2017, pp. 1017–1022.
- [225] T. Höbfeld, L. Skorin-Kapov, P. E. Heegaard, and M. Varela, "Definition of QoE fairness in shared systems," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 184–187, Jan. 2017.
- [226] A. M. Kakhki, F. Li, D. Choffnes, E. Katz-Bassett, and A. Mislove, "BingeOn under the microscope: Understanding t-mobiles zero-rating implementation," in *Proc. Workshop QoE-Based Anal. Manage. Data Commun. Netw.*, 2016, pp. 43–48.
- [227] Google. *Google Global Cache*. Accessed: Dec. 16, 2021. [Online]. Available: <https://peering.google.com/#/options/google-global-cache>
- [228] Netflix. *Netflix Open Connect*. Accessed: Dec. 16, 2021. [Online]. Available: <https://openconnect.netflix.com/en/>
- [229] V. Stocker, G. Smaragdakis, and W. Lehr, "The state of network neutrality regulation," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 50, no. 1, pp. 45–59, Mar. 2020.
- [230] S. Fiegerman. *Net Neutrality Rules are Now Repealed: What it Means*. Accessed: Dec. 16, 2021. [Online]. Available: <https://money.cnn.com/2018/06/11/technology/net-neutrality-repeal-explained/index.html>
- [231] M. M. Maack. *5G is a Threat to Europe's Absolute Net Neutrality*. Accessed: Dec. 16, 2021. [Online]. Available: <https://thenextweb.com/eu/2019/02/28/5g-is-a-threat-to-europes-absolute-net-neutrality/>
- [232] A. Ahmad, L. Atzori, and M. G. Martini, "Qualia: A multilayer solution for QoE passive monitoring at the user terminal," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [233] A. Ahmad, A. Floris, and L. Atzori, "Towards QoE monitoring at user terminal: A monitoring approach based on quality degradation," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2017, pp. 1–6.
- [234] A. Ahmad, A. Floris, and L. Atzori, "Towards information-centric collaborative QoE management using SDN," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
- [235] V. Aggarwal, E. Halepovic, J. Pang, S. Venkataraman, and H. Yan, "Prometheus: Toward quality-of-experience estimation for mobile apps from passive network measurements," in *Proc. 15th Workshop Mobile Comput. Syst. Appl.*, Feb. 2014, pp. 1–6.
- [236] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A system for on-line monitoring of YouTube QoE in operational 3G networks," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 2, pp. 44–46, Dec. 2013.
- [237] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YoMoApp: A tool for analyzing QoE of YouTube HTTP adaptive streaming in mobile networks," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2015, pp. 239–243.
- [238] Q. A. Chen, H. Luo, S. Rosen, Z. M. Mao, K. Iyer, J. Hui, K. Sontineni, and K. Lau, "QoE doctor: Diagnosing mobile app QoE with automated UI control and cross-layer analysis," in *Proc. Conf. Internet Meas. Conf.*, Nov. 2014, pp. 151–164.
- [239] F. Wamser, P. Casas, M. Seufert, C. Moldovan, P. Tran-Gia, and T. Hossfeld, "Modeling the YouTube stack: From packets to quality of experience," *Comput. Netw.*, vol. 109, pp. 211–224, Nov. 2016.

- [240] P. Casas, M. Seufert, F. Wamser, B. Gardlo, A. Sackl, and R. Schatz, "Next to you: Monitoring quality of experience in cellular networks from the end-devices," *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 2, pp. 181–196, Jun. 2016.
- [241] I. Orsolich and L. Skorin-Kapov, "A framework for in-network QoE monitoring of encrypted video streaming," *IEEE Access*, vol. 8, pp. 74691–74706, 2020.
- [242] C. Gutterman, K. Guo, S. Arora, T. Gilliland, X. Wang, L. Wu, E. Katz-Bassett, and G. Zussman, "Requet: Real-time QoE metric detection for encrypted YouTube traffic," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 2, pp. 1–28, 2020.
- [243] S. Wassermann, M. Seufert, P. Casas, L. Gang, and K. Li, "ViCrypt: Real-time, fine-grained prediction of video quality from encrypted streaming traffic," in *Proc. ACM Internet Meas. Conf. (IMC)*, Oct. 2019. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-02375301>
- [244] S. Wassermann, M. Seufert, P. Casas, L. Gang, and K. Li, "ViCrypt to the rescue: Real-time, machine-learning-driven video-QoE monitoring for encrypted streaming traffic," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 4, pp. 2007–2023, Dec. 2020.
- [245] I. Orsolich, D. Pevec, M. Suznjevic, and L. Skorin-Kapov, "A machine learning approach to classifying YouTube QoE based on encrypted network traffic," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 22267–22301, 2017.
- [246] I. Orsolich, L. Skorin-Kapov, and T. Hoffeld, "To share or not to share? How exploitation of context data can improve in-network QoE monitoring of encrypted YouTube streams," in *Proc. 11th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Jun. 2019, pp. 1–3.
- [247] S. Schwarzmann, C. Cassales Marquezan, M. Bosk, H. Liu, R. Trivisonno, and T. Zinner, "Estimating video streaming QoE in the 5G architecture using machine learning," in *Proc. 4th Internet-QoE Workshop QoE-Based Anal. Manage. Data Commun. Netw.*, 2019, pp. 7–12.
- [248] T. Zinner, M. Jarschel, A. Blenk, F. Wamser, and W. Kellerer, "Dynamic application-aware resource management using software-defined networking: Implementation prospects and challenges," in *Proc. IEEE Netw. Oper. Manage. Symp. (NOMS)*, May 2014, pp. 1–6.
- [249] T. Hoßfeld, M. Seufert, C. Sieber, T. Zinner, and P. Tran-Gia, "Identifying QoE optimal adaptation of HTTP adaptive streaming based on subjective studies," *Comput. Netw.*, vol. 81, pp. 320–332, Apr. 2015.
- [250] S. Schwarzmann, C. C. Marquezan, R. Trivisonno, S. Nakajima, and T. Zinner, "Accuracy vs. cost trade-off for machine learning based QoE estimation in 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [251] A. Papa, R. Durner, L. Goratti, T. Rasheed, and W. Kellerer, "Controlling next-generation software-defined RANs," *IEEE Commun. Mag.*, vol. 58, no. 7, pp. 58–64, Jul. 2020.
- [252] A. M. Alba, S. Janardhanan, and W. Kellerer, "Enabling dynamically centralized RAN architectures in 5G and beyond," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 3, pp. 3509–3526, Sep. 2021.
- [253] K. Bouraqia, E. Sabir, M. Sadik, and L. Laidid, "Quality of experience for streaming services: Measurements, challenges and insights," *IEEE Access*, vol. 8, pp. 13341–13361, 2020.
- [254] L. Dinh-Xuan, C. Popp, V. Burger, F. Wamser, and T. Hoßfeld, "Impact of VNF placements on QoE monitoring in the cloud," *Int. J. Netw. Manage.*, vol. 30, no. 3, p. 2053, May 2020.
- [255] K. Tutschku, T. Zinner, A. Nakao, and P. Tran-Gia, "Network virtualization: Implementation steps towards the future internet," in *Proc. Electron. Commun. EASST*, vol. 17, 2009, pp. 1–14.
- [256] E. Grigoriou, A. A. Barakabitze, L. Atzori, L. Sun, and V. Pilloni, "An SDN-approach for QoE management of multimedia services using resource allocation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
- [257] E. Grigoriou, T. Saoulidis, L. Atzori, V. Pilloni, and P. Chatzimisios, "An agent-based QoE monitoring strategy for LTE networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
- [258] A. A. Barakabitze, I.-H. Mkwawa, A. Hines, L. Sun, and E. Ifeachor, "QoEMultiSDN: Management of multimedia services using MPTCP/SR in software defined networks," *IEEE Access*, early access, Nov. 23, 2020, doi: [10.1109/ACCESS.2020.3039953](https://doi.org/10.1109/ACCESS.2020.3039953).
- [259] D. Tsoikas, E. Liotou, N. Passas, and L. Merakos, "A survey on parametric QoE estimation for popular services," *J. Netw. Comput. Appl.*, vol. 77, pp. 1–17, Jan. 2017.
- [260] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.
- [261] T. Hoßfeld, M. Varela, P. E. Heegaard, and L. Skorin-Kapov, "Observations on emerging aspects in QoE modeling and their impact on QoE management," in *Proc. 10th Int. Conf. Quality Multimedia Exp. (QoMEX)*, May 2018, pp. 1–6.
- [262] P. Chen, L. Li, J. Wu, Y. Zhang, and W. Lin, "Temporal reasoning guided QoE evaluation for mobile live video broadcasting," *IEEE Trans. Image Process.*, vol. 30, pp. 3279–3292, 2021.
- [263] F. Bouali, K. Moessner, and M. Fitch, "Multi-timescale QoE provisioning for adaptive video streaming in heterogeneous deployments," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1894–1909, Feb. 2021.
- [264] A. Ahmad, A. Floris, and L. Atzori, "Timber: An SDN based emulation platform for QoE management experimental research," in *Proc. 10th Int. Conf. Quality Multimedia Exp. (QoMEX)*, May 2018, pp. 1–6.



CHAMINDA T. E. R. HEWAGE (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical and information engineering from the University of Ruhuna, Sri Lanka, and the Ph.D. degree from the University of Surrey, U.K., in 2008. From 2004 to 2005, he has worked as a Telecommunication Engineer at Sri Lanka Telecom Ltd. He completed his Ph.D. thesis "Perceptual Quality Driven 3D Video Over Networks" at the Centre for Communication Systems Research (CCSR), University of Surrey. In September 2009, he joined the Wireless and Multimedia Networking Research Group, Kingston University London (KU), London, U.K. He has been with the Cardiff School of Technologies, since 2015. He is currently a Reader with the Cardiff School of Technologies, Cardiff Metropolitan University, and the Lead of the Cybersecurity and Information Networks Centre (CINC). He has worked in number of EU projects namely, FP6 NoE VisnetII, FP7 Optimix, FP7 Concerto, and EU COST Actions QUALINET and 3DCONTOURNET. His research interests include QoE, QoSec, multimedia processing and communications, data visualization, and data security and privacy. He is a member of the IEEE Multimedia Communications Technical Committee.



ARSLAN AHMAD (Member, IEEE) received the engineering degree in aviation electronics (avionics engineering) from the PAF College of Aeronautical Engineering, National University of Sciences and Technology, Pakistan, in 2011, the master's degree in computational sciences and engineering from the Research Center for Modeling and Simulation, National University of Sciences and Technology, in 2014, and the Ph.D. degree in electronics and computer engineering from the Department of Electrical and Electronics Engineering, University of Cagliari, Italy, in 2019. He has worked as a Postdoctoral Researcher with the University of Cagliari (2018–2019). He has worked as a Senior Research and Development Engineer at IS-Wireless, Poland (5G Networks Industry), from 2019 to 2020. He is currently a Senior Lecturer with the Cardiff School of Technologies, Cardiff Metropolitan University, and the Co-Lead of the Cybersecurity and Information Networks Centre (CINC). He has numerous publications in international peer-reviewed conferences and journals, two of which has been awarded the Best Paper Award at IFIP/IEEE IM 2017 and IEEE Multimedia Technical Committee. He has worked as a Marie Curie Fellow (2015–2018) in MSCA ITN QoE-Net. His research interests include multimedia communication, software-defined networking (SDN), network function virtualization (NFV), machine learning, quality of experience (QoE)-based service, and network monitoring and management. The primary focus of his research is the QoE monitoring and management of the over-the-top (OTTs) multimedia services and OTT-ISP collaboration for the joint multimedia service management. He has been the Web Chair of the Tenth International Conference on Quality of Multimedia Experience (QoMEX 2018) and a Technical Program Committee Member of IEEE ICC and IEEE WCNC. He has been a Reviewer of IEEE TRANSACTIONS ON MULTIMEDIA, *Multimedia Tools and Applications* (Springer), and *Signal Processing: Image Communication* (Elsevier).



THANUJA MALLIKARACHCHI (Member, IEEE) received the B.Sc. (Eng.) degree (Hons.) in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2011, and the Ph.D. degree in electronic engineering from the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, U.K., in 2017. From 2011 to 2013, he was a Senior Engineer at Virtusa Private Ltd., Colombo, Sri Lanka. From 2017 to 2018, he was a Research

Fellow in multimedia communication at CVSSP. He is currently a Lecturer in data science and informatics with the Department of Applied Computing and Engineering, Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, U.K., and the Co-Lead of the Cybersecurity and Information Networks Centre (CINC). His research interests include the areas of video coding, processing and communication, computer vision, and machine learning.



NABAJEET BARMAM (Member, IEEE) received the B.Tech. degree in electronics engineering from NIT, Surat, India, the M.Sc. degree in IT from Universität Stuttgart, Germany, and the Ph.D. and M.B.A. degrees from Kingston University London, London. Previously, he was a full-time Lecturer in applied computer science (data science) with Kingston University London. From 2012 to 2015, he worked in different capacities across various industries, including Bell

Labs, Stuttgart, Germany, after which he joined Kingston University London as a Marie Curie Fellow with MSCA ITN QoE-Net, from 2015 to 2018, and a Postdoctoral Research Fellow, from 2019 to 2020. He is currently an Adjunct Lecturer in applied computer science position at Kingston University London. He is a Principal Video Systems Engineer, research at Brightcove, where he works on optimal video encoding strategies, perceptual quality assessment, and end-to-end optimization. He has published in many international conferences and journals and is an active reviewer for many conferences and journals. He is a fellow of the Higher Education Academy (FHEA), U.K. He is a Board Member of the Video Quality Expert Group (VQEG) as a part of the Computer Graphics Imaging (CGI) Group and a member of various standardization groups, such as ITU-T SG16 and MPEG, SVA, CTA, DASH-IF, and UHDf.



MARIA G. MARTINI (Senior Member, IEEE) received the Laurea degree (*summa cum laude*) in electronic engineering from the University of Perugia, Italy, in 1998, and the Ph.D. degree in electronics and computer science from the University of Bologna, Italy, in 2002. She is currently a Professor with the Faculty of Science, Engineering and Computing, Kingston University London, London, U.K., where she also leads the Wireless Multimedia Networking Research

Group and the Course Director of the M.Sc. degree in networking and data communications. She has authored about 200 international scientific articles and book sections, international patents, and contributions to international standards (IEEE and ITU). Her research interests include wireless multimedia networks, video quality assessment, decision theory, machine learning, and medical applications. She has led the KU team in a number of national and international research projects, funded by the European Commission (e.g., OPTIMIX, CONCERTO, QoE-NET, and Qualinet), U.K. research councils (e.g., EPSRC, British Council, and Royal Society), Innovate U.K., and international industries. She is a fellow of the Higher Education Academy (HEA). As an Expert Evaluator and a Panel Member of the European Commission and national funding agencies (e.g., EPSRC in U.K.), she is a part of the NetWorld2020 ETP Expert Group, a Board Member of the Video Quality Expert Group (VQEG), and a member of the IEEE Multimedia Communications Technical Committee, serving in the awards committee, and having served as the Vice-Chair (2014–2016), the Chair for the 3D Rendering, Processing, and the Communications Interest Group (2012–2014), and a Key Member for the QoE and Multimedia Streaming IG. She chairs the IEEE P3333.1.4 standardization working group on the quality assessment of light field imaging. She is an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA (2014–2018) and IEEE Signal Processing Magazine (2018–2021). She was a Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on “QoE-Aware Wireless Multimedia Systems” (2012) and an Editor of IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS (2014), IEEE MULTIMEDIA (2018), *International Journal of Telemedicine and Applications*, and among others.

• • •