

VQEG Klagenfurt 2024 Meeting Minutes

The sessions will be recorded. Let us know if your presentation cannot be recorded, and we will stop the recording. Email presentations to mpinson@ntia.gov, to be posted to the VQEG website.

Monday 1st July 2024

Overview of VQEG Projects

The hosts gave a welcome presentation and information about the logistics
Meeting was started with personal introductions, followed by the status updates from the Working Groups

JEG-Hybrid

#107 Modeling Subject Scoring Behaviors in Subjective Experiments Based on a Discrete Quality Scale

by Lohic Fotio Tiotsop, Politecnico di Torino University, Italy

Novelty: applied to two crowdsourcing datasets, showing how many “problematic” behaviors might be identified (e.g. subject giving almost every time the same score)

This is based on a new approach (probabilistic subject scoring model resulting from the proof of a theorem) to estimate the subjective quality from noisy ratings is proposed and is shown to be more robust to noise than are four state-of-the-art approaches
Also, it makes it possible to highlight several peculiar behaviors typically observed in subjective experiments

Published IEEE TMM 2024 (open access): <https://doi.org/10.1109/TMM.2024.3382483>

#108 Multiple Image Distortion DNN Modeling Individual Subject Quality Assessment

by Lohic Fotio Tiotsop, Politecnico di Torino University, Italy

Training Deep Neural Networks (DNNs) to replicate individual subject assessments of media quality. These DNNs are referred to as Artificial Intelligence-based Observers

(AIOs). Training AIOs is challenging due to the greater noise in individual raw opinion scores compared to the Mean Opinion Score (MOS). We first created a dataset comprising two million samples with synthetic labels, then we trained a customized network for image quality assessment: Multi-Distortion ResNet50 (MDResNet50). Then this network is utilized to initialize the learning process of each AIO. Computational experiments show that:

- i) AIOs well mimic bias and inconsistency
- ii) AIOs exhibit a higher performance with a statistical significance in assessing multiple image distortions
- iii) AIOs more accurately mimic the sensitivity of real subjects to noise and color saturation, and also better predict the opinion score distribution compared to previous AIOs.

AIOs can be downloaded here: <https://media.polito.it/MD-AIOs>

Published ACM TOMM 2024 (open access): <https://doi.org/10.1145/3664198>

#109 Updates on JEG-Hybrid activities (and IGVQM)

by Enrico Masala, Politecnico di Torino University, Italy

Report on the latest activities going on in the JEG-Hybrid group. In particular:

1) new dataset comprising high-resolution images, being prepared with the purpose of training DNN models that would like to predict single user's subjective quality perception. It is expected that this dataset will be partially subjectively annotated in an iterative fashion to improve the performance of the DNN models, in particular the AIOs already developed in the JEG-Hybrid group and presented in past and current VQEG f2f meeting.

2) status of IGVQM project: slowly progressing. Link to the list of considered datasets: https://docs.google.com/spreadsheets/d/1wYVjaqKrSB0z7D2eJLfAd1g6y_831IF7TH6GhAKdKxA/edit?gid=0#gid=0

Objective data computed for AGH-NTIA-Dolby through libVMAF. First results presented (scatter plots etc.). Plan to use standardized format especially for subjective data ([soreal](#) format or [suJSON](#)) and make experiments as reproducible as possible. Plan to make the analysis easy to be re-run to investigate the effect of changing parameters.

#110 Further results on the relationship between SSIM and PSNR for DCT-based compressed images and video

By Maria Martini, Kingston University London

The presentation follows up from the one delivered at a recent VQEG meeting, highlighting the relationship between PSNR and SSIM for DCT-based compressed

images and video. Further results will be presented, including comparisons with other approximations of the relationships between the two.

5GKPI

The session was organized as a workshop on the topic: “Future directions of 5GKPI: Towards 6G”.

#127 Introduction to 5GKPI Workshop (Pablo Pérez & Kjell Brunnström)

Introduction to the workshop motivation and objectives.

The main motivation for the workshop is the information mismatch between CAPs (Content and Application Providers) vs CSPs (Communication Services Providers) and their associated research communities (multimedia vs telecommunications). There is a window of opportunity in the definition of 6G to have an impact on improving this situation.

The long-term goal is develop QoE-QoS models and/or develop a framework to manage QoE in 5G/6G.

The workshop objectives are:

- “Be on the same page”: describe the current status from different perspectives: content vs network, industry vs academia, standardization bodies.
- Identify the key areas to address: how the gap between different parts could be closed.
- Discuss how VQEG could help.

#128 QoS and QoE management in 5G/6G networks (Michele Zorzi)

Keynote given by Michele Zorzi about how QoE is understood and used from the point of view of research in communication technology.

Some key ideas of the presentation:

- Currently use case requirements are defined in terms of QoS metrics (not QoE).
- QoE is a measure of the quality of the data which are transmitted.
 - Example: in LiDAR Point Clouds, QoE is a machine-oriented metric, not a human-oriented measure.

- From a telecommunication perspective, QoE is not so well defined and established as QoS
- Network optimization is balancing the trade-off between QoS and QoE: this can be used for instance as a reward in reinforcement learning approaches
- In random simulations, we accept randomness (e.g. in network behavior). Therefore, the randomness in user behavior is acceptable. We can reduce randomness by modeling users (e.g. learning their preferences)

Concluding remarks:

- Networking and multimedia are more and more intertwined, and will become more so
- Synergistic approaches and joint optimizations are currently being prevented by different (conflicting?) cultures, languages and objectives
- For a successful approach, we need to:
 - Unify (or at least decode/translate) the language
 - Share and understand each other's objectives
 - Understand how technology can support a fulfilling user experience, and how such an experience can be provided based on the available communication performance
- Next steps:
 - Joint discussions and collaborative research
 - Experiments designed to highlight and characterize the key tradeoffs
 - New data sets collected and published with this in mind

#129 Parametric QoE models and QoE management (Tobias Hoßfeld, Pablo Pérez)

Presentation given by Pablo Pérez about fundamentals on QoE management. Key ideas summarized below.

MOS is not enough to understand QoE in systems

- Rating distributions are needed to compute GoB/PoW scores
- Normal model of scores is a good approximation

QoS-to-QoE models are needed

- Simple parametric models can capture the core of the QoS-QoE relationship
- The residual can be predicted with ML models
- This combination provides state-of-the-art QoE estimation (P.1204.3)
- ...but it might not cover your use case

System QoE $E[Q]=E[f(R)]$

- QoE distribution != MOS distribution
- GoB mapping functions are needed ← Capture distribution of subjective scores

#130 Current status of standardization & industry (Kjell Brunnström, Gunilla Berndtsson)

- 3GPP. Many TRs/TSS address aspects related to the problem.
- IETF SCONEPRO: If the companion signaling protocol supports host-to-network metadata, individual packets within a flow can contain metadata describing their drop preference or their reliability
- MPEG. E.g. Per-picture quality metric reporting (SEI) using has been accepted into the VSEI (JVET-AG2032) TuC. Multiple options for both metric and reporting frequency is supported.
- ITU SG-12. Listing of relevant WPs / Questions
- CAMARA QoD API (by the Linux Foundation). The Quality-On-Demand (QoD) API provides programmable interface for developers and other users (capabilities consumers) to request stable latency or throughput managed by Telco networks without the necessity to have an in- depth knowledge of the 4G/5G system or the overall complexity of the Telecom Systems.

A few additional relevant standardization initiatives were identified:

- IEEE HFVE_WG - Human Factors for Visual Experiences Working Group
- Other IETF working groups.
- O-RAN Alliance.

#131 QoE management: Content & Applications Provider (CAP) perspectives (François Blouin)

Presentation on the vision of a CAP (Meta) about QoE management.

Concluding ideas: how can CSPs and CAPs exchange QoE-QoS metrics

1. CAPS agreed on a common metric to be shared
2. An easy-to-understand metric that captures the QoE for a user/viewer of a given service, at the session level (or some other unit of time), at a common numerical scale ([0,100] ?) - on a mobile device

3. Metric needs to be correlated with subjective opinions, with zero-bias and a well-documented accuracy (standard deviation)
4. Metric needs to be either easily calculated by the CSP independently or provided by each CAP via a commonly agreed upon interface (in-band or out-of-band)
5. QoE to QoS correlation needs to be well defined
6. Offer multiple implementations based on existing video quality metrics - ideally, open-source (SSIM, VMAF, FUNQUE, UVQ, Others?)

Metric needs to be actionable, optimization tradeoffs need to be validated and understood.

#132 QoE management: QoE management: Communications Service Provider (CSP) perspectives (Theo Karagioules & Emir Halepovic)

Presentation on the vision of a CSP (AT&T) about QoE management, given by Emir Halepovic.

Concluding ideas: New opportunities open by bidirectional information exchange

- Ability for the network to react to signals and for CAPs to optimize delivery from the source
- What can a network react to?
 - Traffic class
 - Performance requirements (intentional networking)
 - QoE
 - QoS
- Time scale of reactive or pro-active action
 - Real-time:
 - Conditional: QoE or QoS fall outside normal ranges.
 - Pre-session (for the case of QoS or traffic class)
 - In-session / mitigation
 - Post-session
 - Short-term:
 - Expectation of load
 - Event management
 - Long-term:
 - Capacity planning/management

#133 Panel-guided discussion (moderated by Narciso García)

Open discussion on next steps for 5GKPI project.

Panelists:

- Enrico Masala. Statement: we should model user variability.
- Christian Timmerer. Statement: Study usage of CMCD/CMSD in the context of quality (VQEG)

#134 Conclusions (Pablo Pérez)

Potential areas for action.

- Create a common language
 - Currently use case requirements are defined in terms of QoS metrics.
 - From a telecommunication perspective, QoE is a measure of the quality of the data which are transmitted
 - Example of LiDAR Point Clouds QoE is a machine-oriented metric, not a human-oriented measure.
 - Trade-off between QoS and QoE reinforcement learning
- Agree on the right approach to modeling QoS-QoE relationship
 - Which information is needed in each part.
 - Appropriate testing methodology and user modeling
 - Which is the connection between QoS Application KQIs Actual QoE / User experience
 - Which information cannot be shared (privacy, business, etc).
 - Which are the right context factors to take into account
 - Have a correct definition on the metrics
- Define requirements / protocols for applicability
 - How to exchange information between parties
 - Actionable QoE:
 - Real time (resource reservation, react to problems)
 - Post-session: analytics, network dimensioning
- Provide implementation and/or validation
 - Open source implementations
 - Data sets
 - Tests

Next steps

1. Create/update working group Who is interested?
2. Work on white paper (or equivalent)
 - a. Common language
 - b. Right approach to QoE modeling (asking the right questions)

- c. (Discuss before writing! This is a research activity on its own)
- 3. Reach out
 - a. Contribute to SDO
 - b. Develop/test specific models (what use cases?)
 - c. Create data sets

Tuesday 2nd July 2024

QAH

#134 A Dual Rig Approach for Multi-View Video and Spatialized Audio Capture in Medical Training (Joshua MARAVAL)

Sequences are publicly available for scientific research on <https://volumetric-repository.labs.b-com.com/#/immersive-training> , please note that the full sequences will be shared on demand.

HFVE

#104 Comprehensive Insights into Psychological Image Quality: Addressing Ecological Validity and Emotional Content (Kamil and Mikolaj, AGH)

This presentation offers a two-part insight into the realm of image quality assessment. The first part provides an overview of the TUFIQoE project (Towards Better Understanding of Factors Influencing the QoE by More Ecologically-Valid Evaluation Standards) with a focus on challenges related to ecological validity. The second part delves into the 'Psychological Image Quality' experiment, highlighting the influence of emotional content on multimedia quality perception. This approach integrates psychological aspects with technical analysis to enrich Quality of Experience (QoE) assessment, combining diverse disciplines such as psychology, data analytics, and machine learning.

#123 Acceptability & Annoyance in Crowdsourcing (Ali Ak, Nantes University)

VUGC Dataset collected last year used and presented.

Two experiments with same content: AccAnn and ACR-HR
Replication of results in crowdsourcing setting
Focus on acceptability and annoyance categories

Conclusion: transferring the concepts of acceptability and annoyance to crowdsourcing is not straightforward. In crowdsourcing: limited ability to adjust expectations in response to instructions.

Question on instructions and suggestion to have specific separate experiments
Analysis on subject reliability was done and did not show any surprising results.
Crowdsourcing on mobile more complicated due to less control and users could do the tests anywhere - this could impact results.
Prolific platform used.
Content was categorized.

#119 Nantes-MobileHDRVQA Dataset: A Video Quality dataset with Iphone HDR videos and AV1 encoding

(Ali Ak, Nantes University)

Nantes-Mobile HDRVQA Dataset

VMAF was retrained on the dataset

SAM

#106 Pairwise comparisons and JND/JOD quality scale - 8 year perspective

Rafał Mantiuk, University of Cambridge

“In our group, we have been exclusively using the method of pairwise comparisons for the last eight years. We have published multiple datasets, wrote own software for scaling and analysing pairwise comparison data (pwcmp), developed an effective active sampling technique (ASAP), and used pairwise comparisons to make quality scores consistent across datasets. In this talk, I will present lessons learned from that experience - what the main strengths of pairwise comparisons are, what the caveats are, and why we will likely use this protocol for many years to come.”

#114 An Image Quality Dataset with Triplet Comparisons for Multi-dimensional Scaling

Mohsen Jenadeleh, University of Konstanz

“In the early days of perceptual image quality research more than 30 years ago, the multidimensionality of distortions in perceptual space was considered important. However, research focused on scalar quality as measured by mean opinion scores. With our work, we intend to revive interest in this relevant area by presenting a first pilot dataset of annotated triplet comparisons for image quality assessment. It contains one source stimulus together with distorted versions derived from 7 distortion types at 12 levels each. Our crowdsourced and curated dataset contains roughly 50,000 responses to 7,000 triplet comparisons. We show that the multidimensional embedding of the dataset poses a challenge for many established triplet embedding algorithms. Finally, we propose a new reconstruction algorithm, dubbed logistic triplet

embedding (LTE) with Tikhonov regularization. It shows promising performance. This study helps researchers to create larger datasets and better embedding techniques for multidimensional image quality. The dataset includes images and ratings and can be accessed at <https://github.com/jenadeleh/multidimensionalIQA-dataset/tree/main>.”

#115 Impact of feedback on crowdsourced visual quality assessment with paired comparisons

Mohsen Jenadeleh, University of Konstanz

“This work presents a comprehensive investigation into the effects of immediate feedback on crowdworkers’ performance in subjective image quality assessment tasks using paired comparisons. The study is motivated by the need for reliable and efficient crowdsourcing tasks for image quality assessment. A large-scale experiment involving 200 participants was conducted, where participants completed 120 paired comparisons with and without feedback. The feedback informed the workers of the correctness of their responses to comparisons. Almost all of the participants (97%) preferred receiving feedback. The results indicate that feedback reduced response time, improved user experience, and did not cause a bias in the estimation of the just noticeable difference (JND). On the other hand, feedback did not significantly affect accuracy, correlation with the ground truth, or create a learning effect. This study contributes to the field by being one of the first to examine the impact of feedback on crowdworker performance in subjective image quality assessment tasks. The dataset which includes the images and ratings can be accessed at <https://database.mmssp-kn.de/feedback-study-dataset.html>.”

#141 National Differences in Image Quality Assessment: An investigation on three large-scale IQA datasets

Simon H. Del Pin, Norwegian University of Science and Technology

“This paper investigates the potential effects of national differences on image and video quality assessment using discrete rating scales. Drawing on cultural psychology, we hypothesize that observers from different countries may exhibit distinct response styles in interpreting and applying the five-level absolute and degradation category rating scales (ACR, DCR). For our study, we adapt state-of-the-art statistical models for three large-scale image quality datasets (KoniQ-10k, KADID-10k, and NIVD). Our models include country-specific components such as variable rating category thresholds and the probability for extreme ratings on these scales. We found statistically significant differences between ratings collected in different countries. Our results have implications for the analysis and design of current, respectively future datasets and contribute to a more comprehensive understanding of image quality in a global context. We also propose to include lapse rates into statistical models for categorical judgements. Lapse rates model unintentional erroneous responses of subjects in a quality assessment study and provide a regularization mechanism for the scale estimation by maximum likelihood estimation.”

AVHD

#122 Open Source Cloud – reduces the barrier to use and monetize open source

Johan Birmé, Eyevinn Technology

“In this talk Jonas Birmé, VP R&D Eyevinn, will present their mission on lowering the barrier for using open source and by sharing the revenue with the open source creators contribute to a sustainable business for open source. Open Source Cloud offers open source as a service without the creator having to invest in either time or money, removing the need for users of open source to maintain their own infrastructure, and in this session you will learn how that works. Examples of services provided today includes video encoding and quality assurance.” Link to the repository: www.osaas.io

SAM

#141 Subjective Evaluation of Perceptually Optimized Video Encoding Recipes

Andréas Pastor, Nantes Université

“This work presents a new framework for Perceptually Optimized Encoding (POE) in the libaom AV1 codec, which aims to improve perceptual quality and compression efficiency.

The proposed framework uses features extracted from the frame's luma channel to compute a Lagrangian multiplier correction per tube in the RDO process.

The objective evaluation of the framework on the UVG dataset and additional sequences showed promising results with BD-rate savings for some objective quality metrics, but further investigation is needed due to a lack of agreement between metrics.

A pilot subjective evaluation of the proposed perceptual encoding method was conducted on a comprehensive set of SDR-HD video sequences. The results indicate significant bitrate savings and improved perceived quality compared to traditional PSNR and SSIM-based RDO encoding recipes. The methodology used in this evaluation relies on the evaluation of quadruplets performed by a single expert observer, and a statistical analysis method is proposed to conclude on the significance of observed differences.”

#141 Beyond Curves and Thresholds - Introducing Uncertainty Estimation to Satisfied User Ratios for Compressed Video

Hadi Amirpour/Jingwen ZHU

“Just Noticeable Difference (JND) establishes the threshold between two images or videos wherein differences in quality remain imperceptible to an individual. This threshold, collectively known as the Satisfied User Ratio (SUR), holds significant importance in image and video compression applications, ensuring that differences in quality are imperceptible to the majority (p%) of users, known as p%SUR. While substantial efforts have been dedicated to predicting the p%SUR for various encoding parameters (e.g., QP) and quality metrics (e.g., VMAF), referred to as proxies, systematic consideration of the prediction uncertainties associated with these proxies has hitherto remained unexplored. In this paper, we analyze the uncertainty of p%SUR through Confidence Interval (CI) estimation and assess the consistency of various Video Quality Metrics (VQMs) as proxies for SUR. The analysis reveals challenges in directly using p%SUR as ground truth for training models and highlights the need for uncertainty estimation for SUR with different proxies”

ETG

15:30 BST - 16:30 BST

Presentation 116: Assessing Quality of AI Generated Images and Videos,

Abhijay, Portland State University

Abstract: Recent image and video generation models are capable of creating realistic images and videos based on user prompts (text-to-image, image-to-image, and video-to-image). As these new AI models mature, their accuracy and functionality will make them integral to many real-world services and

products. However, before they can be widely integrated and accepted, AI-generated content (AIGC) must be accurate, adhere to the intended usage, and maintain high visual quality to avoid degrading the end user's quality of experience (QoE). Existing traditional image and video quality assessment (IQA/VQA) models focus on fidelity in terms of "reconstruction" quality and are not designed to assess the quality of "generative" artifacts. To address this, newer metrics and models have recently been proposed, but their performance evaluation and suitability have been limited by datasets, performance measures, and applications. In this talk, we shall discuss the current status, gaps, shortcomings, and opportunities that this new paradigm of AI-generated image and video content brings.

Note: This presentation was not recorded and the slides will not be made available.

Presentation 126: Augmented Reality Head-Up displays and digital rear view mirrors in cars

Kjell Brunnström, RISE

Abstract: Another objective was to investigate the impact of different height of cameras and different field of view on the estimation of distance and last safe gap for digital rear view mirrors in cars or trucks.

Q&A: Only feasibility study done, Long term studies into CMS of passenger trucks, driver monitor systems, smart trucks, etc. to follow.

Presentation 137: Energy-aware Spatial and Temporal Resolution Selection for Video Streaming

Mohammad Ghasempour, University of Klagenfurt

Abstract: The growing demand for seamless video delivery and concerns about carbon dioxide emissions have led to the emergence of energy-aware adaptive video streaming. Conventional adaptive streaming approaches, like per-title encoding, only consider the quality when selecting the bitrate ladder for each video content without considering energy consumption. This presentation aims to tackle the challenge of increasing energy consumption in video streaming by introducing a new approach that goes beyond the quality-centric resolution selection approaches. It considers both video quality and decoding energy consumption to construct an optimal bitrate ladder tailored to the unique characteristics of each video content.

QACoViA

Presentation 103: Objective Video Quality Assessment Method for Object Recognition Tasks

Mikołaj Leszczuk, AGH University of Krakow

Abstract: In the field of video quality assessment for object recognition tasks, accurately predicting the impact of different quality factors on recognition algorithms remains a significant challenge. Our study introduces a novel evaluation framework designed to address this gap by focussing on machine vision rather than human perceptual quality metrics. We used advanced machine learning models and custom Video Quality Indicators to enhance the predictive accuracy of object recognition performance under various conditions. Our results indicate a model performance, achieving a mean square error (MSE) of

672.4 and a correlation coefficient of 0.77, which underscores the effectiveness of our approach in real-world scenarios. These findings highlight not only the robustness of our methodology but also its potential applicability in critical areas such as surveillance and telemedicine.

Wednesday 3rd July 2024

AVHD

#102 The effect of viewing distance and display luminance on the visibility of HDR video streaming distortions

Dounia Hammou, University of Cambridge

“While it is well recognized that the visibility of distortions is affected by the viewing distance and display peak luminance, very few datasets control those conditions, and also few video quality metrics can account for them. To address this gap, we collected a new video quality dataset, HDR-VDC, which captures the quality degradation of HDR content due to AV1 coding artifacts and the resolution reduction. The quality drop was measured at two viewing distances, corresponding to 60 and 120 pixels per visual degree, and two display mean luminance levels, 51 and 5.6 nits. In contrast to the existing datasets that use direct rating protocol, we employ a highly sensitive pairwise comparison protocol with active sampling and comparisons across viewing distances to ensure possibly accurate quality measurements. We also provide the first publicly available dataset that measures the effect of display peak luminance and includes HDR videos encoded with AV1. Our results indicate that the effect of both viewing distance and display luminance is significant, and it reduces the visibility of coding and upsampling artifacts on dimmer displays or those seen from a further distance. The dataset is available at <https://doi.org/10.17863/CAM.107964> and the code at <https://github.com/gfxdisp/HDR-VDC>.”

#112 Stimulus presentation structure and order - impact on subjective assessment

Tomasz Konaszyński, AGH University of Krakow

“Presentation concerns a study of video quality perceived by the user (Quality of Experience, QoE) that was conducted, with an examination whether and how the order and structure of the video sequence presentation affects subjective assessment. For this purpose, an influence of content variability/repeatability, quality of the preceding sequence and sequence order were analyzed. Observations on the correlation of QoE with the micro-structure of sequence presentation are described, which can be the basis for hypotheses for dependence of QoE assessment on the above-mentioned factors. Observed relationships regarding influence of the number of video repetitions and impact of the predecessor's quality on subjective evaluation are consistent with research work on the influence of order/arrangement and structure of research stimuli on results of subjective evaluations.”

#117 AVT-VQDB-UHD-2-HDR: An open 8K HDR source dataset for video quality research

Dominik Keller, Technische Universität Ilmenau

“The market is seeing an increasing availability of screens with ultra-high definition (UHD) and High Dynamic Range (HDR). Additionally, video streaming platforms like YouTube are already offering HDR content with resolutions up to 8K (UHD-2), and the volume of such material is expected to grow. Consequently, manufacturers and service providers need robust video quality models to assess their

customers' Quality of Experience. Currently, the most widely used video quality models support resolutions only up to 4K (UHD-1). However, with the rise of 8K resolution content, accurate video quality estimation for these higher resolutions is becoming crucial. In this paper, we introduce the AVT-VQDB-UHD-2-HDR dataset, which comprises 31 8K HDR video sources, each 15 seconds long. These videos were created to accurately represent real-life footage while addressing video coding and quality testing challenges. We provide a detailed explanation of the methodologies used in planning, creating, and post-processing these video sources. Additionally, we evaluate aspects such as detail, dynamic range, and the suitability of the content for video coding research. The videos and objective descriptors are publicly available for future research at <https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-2-HDR>

#124 QoE Based Objective Analysis of Low-Latency Algorithms in DASH.JS

Syed Uddin, AGH University of Krakow

“In the past decade, significant growth in video traffic has presented many challenges for live video streaming providers to deliver to ensure consistent quality of experience (QoE) for viewers. Low latency live streaming over HTTP using Dynamic Adaptive Streaming over HTTP (LL-DASH) has emerged to deliver high quality video content and short end-to-end latency. Multiple low-latency based adaptive algorithms have been proposed such as Learn2Adapt (L2A-LL) and Low on Latency (LOL+), which makes decisions based on heuristic predictive models and learning models. This paper analyzes how effectively low-latency algorithms enhance the user experience. The low-latency algorithms are compared with traditional HTTP Adaptive Streaming (HAS) algorithms. The DASH.js environment is used for experimental evaluation. The experimental results demonstrate that the L2A-LL algorithm achieves the highest bitrate, followed by the LOL+ algorithm. The low-latency algorithm achieves a higher video rate compared to the standard adaptive streaming algorithms. The findings will help identify the limitations of existing adaptive algorithms and inform strategies for their enhancement.”

#121 Crowdsourcing Subjective Video Quality Assessment

Avrajyoti Dutta, AGH University of Krakow

“This study investigates the evaluation of subjective video quality utilizing short video clips on a crowd-sourcing platform. A total of 47 participants provided ratings for 170 short videos using the Absolute Category Rating (ACR) scale. Although all short videos are in full HD, clips of fair quality resulted in lengthier response times (1867.19 ms), exceeding other categories within a 5000 ms limit. The experiment was only conducted on Google Chrome in incognito mode to eliminate extraneous interference. The use of automated sequential playing allowed for efficient data collection, resulting in the completion of the experiment in around 55 minutes. The results indicate that videos of Fair quality regularly received higher timings, suggesting the presence of potential biases in subjective evaluations. The study emphasizes the significance of taking into account contextual variables, such as monitor resolution and experimental design, when evaluating video quality. The findings improve our understanding of how users perceive video quality and provide valuable insights for future research on multimedia technology.”

IMG

#111 KULF-TT53: A Display-Specific Turntable-Based Light Field Dataset for Subjective Quality Assessment

The presentation will introduce a novel display-specific light field dataset, captured using a DSLR camera and a turntable rig. The visual data of the seven static scenes were recorded twice by using two settings of angular resolution. While both were acquired uniformly within a 53-degree angle, which matches the viewing cone of the display they were captured for, one dataset consists of 70 views per content, while the other of 140. Capturing the contents twice was a more straightforward solution than downsampling, as the latter approach could either degrade the quality or make the FOV size inaccurate. The paper provides a detailed characterization of the captured contents, as well as compressed variations of the contents with various codecs, together with the calculated values of commonly-used quality metrics for the compressed light field contents. We expect that this dataset will be useful for the research community working on light field compression, processing, and quality assessment, for instance to perform subjective quality assessment tests on a display with a 53-degree display cone and to test new interpolation methods and objective quality metrics. In future work, we will also focus on subjective tests and provide relevant results. This dataset is free to access for the research community.

#125 A Light-Field Video Dataset of Scenes with Moving Objects Captured with a Plenoptic Video Camera

Light field dataset obtained with Raytrix R8 plenoptic camera, including compressed versions of the videos, objective quality scores, and content characterisation. NOTE: the presentation can be merged with the previous one (ID 111) if required.

#118 Status of IMG test plan on XR communication systems for P.IXC

Current status of the test plan. Description of the tests being performed in the 13 different labs. Agreed to have an IMG meeting at CWI from 2-4 September 2024

#135 Subjective Evaluation of the Impact of Spatial Audio on Triadic Communication in Virtual Reality

Virtual Reality (VR) enables users to meet, converse, and collaborate in shared virtual environments. For such communication systems, many system factors can affect user experience and perception. To effectively allocate system resources, understanding of the relative influence of such factors is required. One important factor is a spatial auralization, which has been shown to elevate users' experience in traditional and single-user VR systems. However, its effect in multi-party social VR has not been fully investigated. In this work, we conducted a study in context of the VQEG-IMG joint test plan assessing the effect of spatial audio on audiovisual plausibility and presence perception in a three-user interactive communication scenario. Triads of participants perform a collaborative conversation task under three conditions: a VR condition with binaural spatial audio, a VR condition with simple diotic audio, and a real-world reference condition. This paper presents the results of the study based on questionnaire-based evaluation.

#120 Subjective evaluation of immersive communication systems: A case study on the impact of display technology with a Free Viewpoint Video system

Presentation of one of the experiments that have been performed within the IMG joint test plan and ITU-T P.IXC . In particular, this experiment addressed the subjective evaluation of the impact of display technology in a remote communication scenario with a real-time Free Viewpoint Video system

#140 AVT-ECClass-VR: An open-source audiovisual 360° video and immersive CGI multi-talker dataset to evaluate cognitive performance

The paper is part of a project to assess how complex visual and acoustic scenes affect cognitive performance in classroom scenarios, across age groups from children to adults. Here, the potential of audiovisual virtual environments for systematic user studies is explored. As of now, most studies have examined rather simple acoustic and visual representations, which do not reflect the reality of school children. An adapted version of the audiovisual scene analysis paradigm introduced by Ahrens et al. is presented, focusing on the localization and identification of talkers within a scene. The dataset includes two audiovisual scenarios (360° video and computer-generated imagery) and two implementations for dataset playback. The paper details the recording and post-processing of the content. The 360° video part of the dataset features 200 video and singlechannel audio recordings of 20 speakers reading ten stories, and 20 videos of speakers in silence, resulting in a total of 220 video and 200 audio recordings. The dataset also includes one 360° background image of a real primary school classroom scene, targeting young school children for subsequent subjective tests. All stories were recorded in the German language with native German speakers. The second part of the dataset comprises 20 different 3D models of the speakers and a computer-generated classroom scene, along with an immersive audiovisual virtual environment implementation that can be interacted with using an HTC Vive controller. Both implementations also include a Unity plugin to connect and interact with the Virtual Acoustics auralization software. As a proof of concept, the dataset includes example output data collected from ongoing perception tests. There, subjects have the task of identifying which talker in the scene is reading out which story, using the story-to-speaker mapping input system developed within this paper.

Link to dataset:

<https://github.com/Telecommunication-Telemedia-Assessment/AVT-ECClass-VR>

VQEG Admin session

Autumn On-line meeting: host? AGH? Univ of Aachen RWTH?

Competing event: VCIP, Dec 8-11;

Spring hybrid meeting: host Meta? April May

Competing event:?

Action point (Kjell): Audio call for the board to decide on upcoming meetings and new WG on user behavior and HFVE

Action point (Mikolaj): Draft a proposal for new WG

NORM

#146 Detection of Hard-To-Compress Video Sequences Based on Objective Quality Metrics

Filip Korus, AGH

Video streaming services account for 82% of all IP traffic (as of 2022, according to Cisco). Netflix, TikTok, YouTube and Twitch are gaining popularity, attracting the attention of both fans of films and series, as well as computer game enthusiasts. However, compressing video content generated during esports events becomes a challenge, even for advanced video codecs. Research based on video quality metrics reveals these difficulties. As a result, a machine learning model was developed to assess compression difficulty.

#144 From VCA to DeepVCA and Grand Challenge on Video Complexity

Hadi Amirpour, University of Klagenfurt

“A summary of activities in video complexity analysis”

#145 A dataset for understanding open UGC datasets

Pierre Lebreton, CAPACITES, Nantes Université

“User Generated Content (UGC) video streaming is a major application on the Internet. Even small bitrate savings can have large network impacts at this scale. In order to achieve improvements without sacrificing experience, the quality of UGC videos needs to be better understood. In recent years video quality evaluation models designed for the evaluation of UGC videos have received a lot of attention. However, considering that these models are learning-based models, they heavily depend on the training data that has been used. In this paper, a new dataset is introduced that allows studying the differences in characteristics between existing UGC video datasets. It reveals the range of quality that was covered by existing UGC video datasets, and the implication of these quality ranges on training and validation performance of UGC video quality prediction models. Furthermore, this work demonstrates that dataset alignment enables existing UGC models to achieve higher performance. This alignment dataset can be found openly available on Zenodo (<https://zenodo.org/doi/10.5281/zenodo.12155934>).”

ETG

#142 Updates on testing activities in MPEG VQA

Mathias Wien, RWTH Aachen University

This presentation reports on recent and upcoming visual testing activities in ISO/IEC JTC 1/SC 29/AG 5 MPEG Visual Quality Assessment. Besides work in the context of starting standardization projects and verification tests, AG 5 has started work for building the CVQM database (Coded Video for study of Quality Metrics) including conventional and neural network-based video coding schemes. The goal is to evaluate objective quality metrics for their suitability for decision making in the standardization process on test material with potentially previously unseen impairments.

NORM

#113 COVER: A Comprehensive Video Quality Evaluator

Zhengzhong Tu, Texas A&M University

“Video quality assessment, especially for a massive scale of user-generated content, is an essential yet challenging computer vision and video analysis problem. Prior methods have been shown to be effective in mirroring subjective human opinion scores; however, they fail to capture the complicated,

multi-dimensional aspects of factors that impact the overall perceptual quality. In this paper, we introduce COVER, a comprehensive video quality evaluator, a novel framework designed to evaluate video quality holistically --- from a technical, aesthetic, and semantic perspective. Specifically, COVER leverages three parallel branches: (1) a Swin Transformer backbone implemented on spatially sampled crops to predict technical quality; (2) a ConvNet employed on subsampled frames to derive aesthetic quality; (3) a CLIP image encoder executed on resized frames to obtain semantic quality. We further propose a simplified cross-gating block to interact with the three branches before feeding into the predicting head. The final quality score is attained using a weighted sum of each sub-score, making a multi-faceted metric. Our experimental results demonstrate that COVER exceeds the state-of-the-art models in multiple UGC video quality datasets. Moreover, COVER offers a diagnosable quality report to explain the quality score in multiple pillars, while it is capable of processing 1080p videos at 3x faster speed than the real-time requirement. To facilitate future research on efficient and explainable video quality research, the code is available at <https://github.com/vztu/COVER>.”

ETG

#148 Subjective QoE Assessment for Virtual Reality Cloud-based First-Person Shooter Game

Henrique Souza Rossi, (LTU) Luleå University of Technology

The recent move of the cloud gaming (CG) industry to virtual reality (VR) platforms brings the benefits of the cloud to the most immersive quality of service (QoS) and QoE-sensitive VR content. Virtual reality cloud-based gaming (VRCG) necessitates understanding of stochastic broadband network connections on users' QoE so that stakeholders can deliver quality content by optimizing their services to underlying QoS conditions. Very few studies exist in the literature that study the impact of network QoS on users' QoE for VRCG. This paper presents subjective tests (N= 30) and investigates the effect of network-emulated QoS metrics (N= 28) on the commercial Nvidia CloudXR service and their impact on the users' perceived QoE while playing Serious Sam VR shooter game. Our findings reveal that QoE was most affected by round trip time (RTT) ≥ 75 ms or packet loss (PL) $> 6\%$. Random jitter (RJ) caused QoE degradation for values more significant than one standard deviation, while the combined RTT and PL degraded QoE the most for RTT ≥ 25 ms and PL $\geq 4\%$. Finally, based on actual network traffic data between Sweden and various data centers in Europe, we suggest VRCG can be hosted anywhere in these data centers with minimal impact on QoE for wired connections. However, for 4G and 5G high jitter values could pose a challenge to VRCG services.

<https://github.com/hsr-research/ALTRUIST>

Thursday 4th July 2024 and Friday 5th July 2024

IRG-AVQA

Chaired by Chulhee Lee and Alexander Raake

Chulhee report on progress in ITU-R WP6C

See slides from Chulhee Lee and Alexander Raake more information and notes

To join meetings of Wednesday meeting of P.BBQCG contact David Lindero

(david.lindero@ericsson.com), Alexander Raake (alexander.raake@tu-ilmenau.de) or Martin

Adolph (Martin.adolph@itu.int)

<https://www.itu.int/rec/T-REC-G.1052-202405-P>

ITU-T SG 12 Q19 Interim meeting

See slides from Chulhee Lee more information and notes

Meeting adjourned 09:45