VQEG Meeting

Brightcove, UK

December 12 to 16, 2022

Video Meeting Recordings are available here: <https://drive.google.com/drive/folders/1vEDujX_THFEW6fLBtF1IXJdV2AKAj6Ro?usp=sharing>

Reminder: send your slides to Margaret ([mpinson@ntia.gov](mailto:mpinson@ntia.gov)) with the presentation number!

Day 1 (Monday, December 12)

Overview of VQEG projects

See the video recording for details.

QACoViA Session

#240 QACoVia Overview, by Mikołaj Leszczuk of AGH

See the video recording or slides for details.

#206 “Method for Assessing Objective Video Quality for Automatic License Plate Recognition Tasks” by Mikołaj Leszczuk of AGH

Proposes an objective model for evaluating video quality for automatic license plate recognition. Studied precision and recall, compared to a large variety of simulated camera capture impairment (e.g., noise, JPEG compression, blur).

The paper associated with this presentation can be found here: Leszczuk, M., Janowski, L., Nawała, J., Boev, A. (2022). Method for Assessing Objective Video Quality for Automatic License Plate Recognition Tasks. In: Dziech, A., Mees, W., Niemiec, M. (eds) Multimedia Communications, Services and Security. MCSS 2022. Communications in Computer and Information Science, vol 1689. Springer, Cham. <https://doi.org/10.1007/978-3-031-20215-5_13>

Comment from Ioannis: link with PCS, “Neural compression” session (last day): [7-9 December 2022 • San Jose, California, USA](https://2022.picturecodingsymposium.org/view_event.php?mid=3)

#207 “Assessing Rail 8KUHD CCTV Facing Video” by Femi Adeyemi-Ejeye of University of Surrey, UK

*Note:* This presentation could not be recorded.

Studying the usability of 8K forward facing cameras on trains, for investigation of collisions. Jerky motion was less objectionable, because the critical factor was being able to see details in still frames. Subjects were rail experts, including rail operators and police. People with 15+ years of experience had better recognition rates; they also gave higher MOSs. This had a major impact on the bitrate needed for recognition.

Question from Ioannis: how were the results for 10 and 20 Mbps compared to 30Mbps?

Answer: the 30Mbps streams all received the highest score (4) from all experts

Question from Vittorio (who also had problems with connection): which kind of display / which device to play the videos? → 8K Dell monitor. Did you try the same experiment using 4K content instead of 8K? → Thinking about it now. Higher frame rate.

Question from Margaret: are you going to publish the results? Interested in the impact of experience. Hard to find studies. → Currently working on a paper.

Note about the dataset: negotiating to have it open source.

#214 “Comparing the Robustness of Humans and Deep Neural Networks on Facial Expression Recognition” by Lucie Lévêque of Nantes Université, France

The paper associated to this presentation can be found here: Lévêque, L.; Villoteau, F.; Sampaio, E.V.B.; Perreira Da Silva, M.; Le Callet, P. Comparing the Robustness of Humans and Deep Neural Networks on Facial Expression Recognition. *Electronics* **2022**, *11*, 4030. <https://doi.org/10.3390/electronics11234030>

**ACM International conference on Multimedia Experience (IMX)** happening in Nantes (France) next year (June 13-15, 2023)! More details here: [ACM IMX 2023](https://imx.acm.org/2023/)

* Deadline for workshop proposals: December 15
* Call for associate chairs: January 2
* Call for technical papers: February 3

Margaret Pinson asked about confidence levels and how to implement facial recognition with multiple answer options (e.g., fear and disgust).

Margaret Pinson asked if subjects can select more than one emotion during the subjective experiment.

Margaret Pinson asked about a couple of strong vertical lines. Answer: neutral emotion was the default for this algorithm.

#242 “Video Coding for Machines: Large-Scale Evaluation of DNNs Robustness to Compression Artifacts for Semantic Segmentation” by Alban Marie, Karol Desnos, Luce Morin, Lu Zhang of University of Rennes, France

(Pre-recorded) This came from a recent conference poster session.

Conference was **MMSP 2022**, and the poster is available in the VQEG meeting files.

CGI

#225 “Short updates on ITU-T P.BBQCG” from Saman Zadtootaghaj of SIE

ITU-T work item P.BBQCG, to develop a gaming QoE model that uses the bitstream meta data for video quality module. See the presentation slides for details.

Interactive test and passive test (watch only) plans presented. They are **seeking labs who would be willing to run the subjective tests**; these labs will be given access to all of the datasets for research purposes.

Compression will be applied as post-processing (i.e., will not change rendering decisions).

Participation requires signing a **multiple-party NDA**. This must be signed as-is (i.e., the text cannot be modified). Contact Saman if you are interested in participating.

Question: Could individual subject ratings be shared without any information on media files? This would aid research into subject screening and the accuracy of subjective testing. Saman will look into this question.

#230 “QoE in 3GPP” from Pablo Pérez of Nokia

We described 3GPP Recommendations that relate to QoE, particularly QoE metric collection. See the slides for details.

**The meeting minutes have been approved to this point.**

Day 2 (Tuesday, December 13)

NORM

#213 “Summary of live per-title encoding approaches using video complexity features” by Vignesh V Menon of University of Klagenfurt

NR metric to enable live streaming with perceptually aware variable bitrate encoding. Slides will be distributed (<https://vqeg.org/VQEGSharedFiles/MeetingFiles/2022_12_Brightcove_UK/VQEG_NORM_2022_213.pdf> ). This is a continuation of content adaptive encoding, used for out of service streaming.

**VMAF 6, Ioannis recommends the following dataset for training metrics**: <<H. Wang, I. Katsavounidis, J. Zhou, J. Park, S. Lei, X. Zhou, M.-O. Pun, X. Jin, R. Wang, X. Wang et al., “VideoSet: A large-scale compressed video quality dataset based on JND measurement,” Journal of Visual Communication and Image Representation, no. 46, pp. 292–302, 2017.>> The dataset (“VideoSet”) itself is hosted on IEEE: <https://ieee-dataport.org/documents/videoset>

Ioannis recommends comparing with the exact same number of encodings.

Be cautious of the hidden assumption that can achieve interpolated bitrate/quality encodings, which may not be true. Staircase function.

For relative energy differences, the presence of hardware encoders can change the statistics.

Can use proxy (like H.264, which runs 5x faster than H.265) to predict target encoding curve. Energy component mentioned in this presentation is very important.

#205 “Video Metadata” by Ioannis Katsavounidis of Meta

Idea is to insert metadata into video stream with information on video metrics. Idea is mature, so VQEG approached MPEG. Asked for documents to be reviewed and considered by MPEG. The response was that because VQEG does not have a legal presence, MPEG could not accept this document. MPEG requested that a member of MPEG make a Contribution on this topic. Ioannis (as a member of MPEG) will take this action.

A similar memo was sent to AOMedia. The goal was to establish a relationship with AOMedia, which would let multiple VQEG groups communicate their efforts to AOMedia (e.g., best practices on subjective testing, IMG, and NORM). AOMedia replied with a liaison agreement that someone at VQEG needs to sign.

#216 “Motion search” by Ioannis Katsavounidis, Cosmin Stejerean of Meta

<https://vqeg.org/VQEGSharedFiles/MeetingFiles/2022_12_Brightcove_UK/VQEG_NORM_2022_205_%20motion_search.pptx>

Goal is to order videos by coding complexity. Diagram of generic encoder. Software calculates for each frame:

* the number of intra-coded 16x16 blocks
* number of P-coded blocks
* number of B-coded blocks
* error (residual energy after subtract)
* bits (estimate of bits to encode the frame)

Software is open source, available on GitHub, in C++. <https://github.com/facebookresearch/motion-search.git>

Which mode to pick for NR metric development, encoding complexity estimation?

* IPPP if videoconferencing
* I if image only
* IBBP if video on demand
* If must standardize, IPPP is the most natural. Runs faster. Fundamental
* Can detect scene cuts by increase in number of I frames

#217 “Discussion on supplementing classic SI and TI with improved complexity metric (VCA, motion search, etc.)” by Ioannis Katsavounidis of Meta

Goal is to retain SI/TI, but also develop improved metrics.

Lukas: target use case?

Ioannis: Given a long video sequence, deciding **where to spend more bits** (rate control). Produce feature (vectors) instead of a single number. Don’t want hundreds of features; need explainable features that capture fundamental concepts. Energy, entropy, etc. Then fuse into an estimator of coding complexity. This will make it future proof. (a) computationally efficient, (b) meaningful, (c) paired with a simple fusion model.

Lukas: how to define complexity? This will help guide the research. Agree on explainability.

Cosmin: working on SI/TI, struggles predicting the **compressability** of video. We know this is a gap. **Understand diversity of dataset** is also useful (e.g., what is redundant, not helpful).

Lukas: (d) agnostic, not tied to one single codec.

Ioannis: some videos switch complexity when change AVC to HEVC. May need multiple fusion models. Let’s fix a baseline (conservative) 16x16 blocks, no complex block partitions

Margaret: use case for managed networks. 3GPP wants to understand better the input video that is being compressed, to optimize the future bandwidth use. **Priority access**. Should the managed network increase or decrease allocation to a particular video stream, by understanding the impact on the end user. Must be forward looking.

Lucjan: also very interested in this use case. Understand video content.

Lukas: camera impact on quality vs compressibility. Content dependance. Potential impact of future events (unknown). People vs computer vision observer will have an impact, too.

Ioannis: Motion search, embedded into video stream, can help with prioritization and making the network side easier.

Lucjan: 1st use case is not just camera features but slight prediction will

Margaret: use case dependency exists, can likely trust future networks to help us detect the use case.

Ioannis: motion vector can detect camera jiggle, zoom in, zoom out.

Lucjan: optimization of 5G, where network asks you to limit your resources. Theoretical work. Balance multiple devices with limited resources. Factory, automatic system / computer vision. Could be used for archiving (don’t know which will be useful)

Nabajeet: input is different modes, plus amount of information available, produce complexity. Can still get assessment with little information, but estimation improves as more information becomes available.

Cosmin: may be hard to be truly codec agnostic. Want to understand situations where this is likely to occur. **Are there features we can extract, that will predict these outliers?** Characteristics? (e.g., based on motion vectors).

Ioannis: single feature correlates best seems to be block size (16x16 vs 8x8 vs 32x32), small fragmented motion is better handled by newer codecs. This is based on past experience.

Cosmin: Motivation. If have limited compute budget, can’t run them all, must prioritize which codec is used, which videos are run through multiple codecs. Important videos get more advanced codecs. This metric could identify videos that will benefit from different codecs. Don’t waste energy. Where is it worth to invest energy.

#203 “New datasets: camera noise and compression” by Margaret Pinson of NTIA/ITS

Ioannis: supports the CRF factors chosen for the VCRDCI dataset.

Lucjan: take 5 to 10 pictures in each camera and lighting situation. Or use an artificial support to eliminate hand shake.. Hand shake in low light is a major variable, so we want to make sure that it is equally represented across all lighting situations and cameras.

(After the break) #235 “Cross-Resolution Image Quality Assessment” by Oliver Wiedemann of University of Konstanz

Motivation: users up-scale and down-scale images on mobile devices. Change in resolution impacts attention, but prior datasets only include each image at a single resolution. Large dataset with 2 annotations per subject (420 images x 3 resolutions). Expert subjects. Higher agreement among subjects on 4K resolution than SD resolution. Dataset is available to researchers.

Ioannis: viewing distance? Oliver: fixed viewing distance requested. Ioannis: Meta uses “FBMOS” (Facebook MOS), similar to proposal here. Viewport adaptive score (e.g., 1080p rating, vs scale original and degraded down to 720p and compute rating).

Margaret: confirmed that each subject rated each image twice. Oliver: batches of 25 images, then repeated, if 2nd set did not agree within 0.9 correlation, then had to score the set again. Lucjan: did people “cheat” the system? Oliver: no evidence of cheating on the plots. Motivation was to detect fatigue. Vlad: plotted scatter plots to establish minimum threshold requirement, saw steady pace above this threshold, saw one cheating but it was very obvious (e.g., used fewer options). Ioannis: consider using the new MLE subject screening. Interested in insights, on inconsistencies from particu

<https://arxiv.org/abs/2212.05813>

#202 “Why No Reference Metrics for Image and Video Quality Lack Accuracy and Reproducibility” by Margaret Pinson of NTIA/ITS

* We have journal paper [related to the talk](https://ieeexplore.ieee.org/document/9837932)
* Metrics 0.96-1.0 -> overtrained
* Metrics 0.9-0.96 is extraordinary -> we need extraordinary proofs :)
* Aesthetics in influencing our results

Lucjan: Maybe we should train subjects? We need research in this direction. I am open to joining forces on this topic.

* Analysis of literature shows “accuracy claim” without real evidence!
* Single dataset strategy cannot be justified

Kjell: Question about the graphs. Order is just to make it easier to read

Ioannis: It is often when people use metrics wrongly. Ioannis showed that some people used it wrongly, can we estimate how frequently it is the case?

Margaret: Cannot estimate it. Most of the time, there is no information in the paper about how a metric was used. We cannot trust evaluations, for the same datasets we have different values.

* There is Margaret’s github repository where implementation of different metrics is available

Ioannis: 3x baseline - means 3 times slower than baseline

Lucjan: If a metric detects blur, can we claim it is not working on those datasets?

Margaret: Yes, the main impairment is blur.

Francois: We have very different evaluations compared to the authors? -> yes, authors are overoptimistic

Ioannis: Part of the blame is us. If you do not have 0.95 you do not get publication, so we see this over optimistic results.

Lucjan: We have tool to reject papers which are over optimistic

Enrico: The results are based on manual investigation? -> not all the results

Narciso: The presentation was focused on NR, we concern about databases. Can you link databases which we can trust? -> Next step is to evaluate datasets

* Different databases are different in many ways, some metrics/features can be contradicting for the different cases

Ioannis: Thank you for great work! Discussion about the dataset is needed. We have to collect datasets which are better/worse, maybe by testing numerous metrics on a dataset and conclude if this dataset is ok.

Lucjan: With a large dataset we obviously observe a strong scatter of the results.

Francois: What is the main idea of RCA? -> How to combine scores from different metrics to overall MOS. By changing the combination function, we can get results for other use cases.

Ioannis: Extension to NR which was used for VMAF is probably the best approach.

* The raw data are available for people interested

Repository: <https://github.com/NTIA/NRMetricFramework>

Discussion on BT.500 method, Annex 8 to Part 2, “Expert viewing protocol to video material”

11-level numerical scale; with impairment scale (imperceptible, slightly perceptible, perceptible, clearly perceptible, annoying and slightly annoying); and with categories (everywhere and somewhere) as sub-categories for levels 0 to 9 (slightly perceptible through slightly annoying)

Ioannis: this combines two separate evaluations into one, being used by standards developing organizations. Hard to justify the ranking of the (everywhere and somewhere) with the numbers (e.g., 3 vs 4). Next meeting is May 2023

Lucjan: agree this is a serious problem. How was this analyzed? How do people react? Would be very interested in seeing results of such a study, if it exists.

Ioannis: subjects using the scale is one issue. Non-expert

Proposal: send ITU statement of concern with this method, challenge method, concise and simple letter. Capture key concerns.

Propose removing from BT.500,  unless there are studies conducted of which we are unaware, made available publicly, supporting this method. (doubt our knowledge)

Oppose adding additional words to BT.500. If we don’t have proof it works, then the mis-use of BT.500 will occur. It should only contain methods that are proven to work, with some care. Better to remove. If experiments are run, we should try to do analysis for the community (i.e., both our concerns and conclusions reached by analyses comparing with other methods).

* Lucjan, Ioannis, Narciso Kjell, Enrico
* Draft liaison statement to be considered on Wednesday

**The minutes were approved to this point on Tuesday.**

Day 3 (Wednesday, December 14)

#208 “Assessing video content complexity” by Maria Martini of Kingston University London

Compare existing and proposed metrics in terms of their ability to assess the compressability of videos. 12 proposed complexity metrics are listed. Compare with area under the curve that is formed by the drop of quality in response to bitrate. Considered possible combinations, to estimate overall complexity from separate measures of spatial and temporal complexity.

Computed “area under the curve” for the same range of bitrates, for each media.

Discussion on motion search for complexity estimation.

Miscellaneous Topics

Late arrivals that could not be collocated with similar topics

#236 “D-JNDQ: From Just Noticeable Differences to (HDR&SDR) Image (and Video) Quality” by Ali Ak of Nantes Université

Presents a model based on JND for image and video quality both for HDR and SDR.

<https://github.com/kyillene/D-JNDQ>

#227 “Efficients Deep-Based Graph Metric For Point Cloud Quality Assessment” by Marouane Tliba of Université d'Orleans

Presents a objective quality metric for point cloud based on machine learning

#239 “Perceptual quality of video on simulated low temperatures in LCD vehicle displays” by Kjell Brunnström of RISE Research Institutes of Sweden

A presentation about the impact on low temperatures on LCD displays and the acceptability of the distortions on them in relation to car driving. Very low temperatures cause the LCD display to blur motion, which makes it more difficult to see moving vehicles.

#244 “On the benefit of parameter-driven approaches for the modeling and the prediction of Satisfied User Ratio for compressed videos” by Jingwen ZHU of Nantes Université

A presentation about predicting the Satisfied User Ratio (SUR) in videos

Administrative Discussions

Next meeting: Sony potential host (Saman) try to host in the US, best would be week before or after CVPR: **June 12 to 16 or June 26 to 30. Location: headquarters at San Mateo, CA, or San Diego, CA.**

Adjacent events:

* QoMEX: June 20-22, 2023 - Ghent, Belgium
* ACM IMX: June 13-15, 2023 - Nantes, France

VQEG poll will be sent to the main reflector, after Saman consults logistics.

Proposal for a new VQEG Group

Exploratory group: **Future/Emerging Technologies**. Proposed by Nabajeet Barman and Saman Zadtootaghaj. Proposed group description is here: <https://docs.google.com/document/d/164EQ02T3KtPmcZQf3_pR1AXIz4Vryd0N--jiPKilku0/edit>

Propose to hold monthly calls.

Recommend choosing a specific topic to start (e.g., AI-based technologies: super resolution and learning based video compression).

**Agreed to create the group.** Discussions on scope will continue on Friday. Nabajeet and Saman will propose an updated scope that does not overlap other VQEG groups.

IMG

#234 “Video Quality Assessment based on Quality Aggregation Networks” by Yaosi Hu, Zhenzhong Chen of Wuhan University

Try to simulate human perceptual effects. Short-term effect: visual masking effect. Long-term effect: memory effect. Describe metric named Quality Aggregation Network (QAN) and compare the performance of QAN to PSNR and other metrics.

#209 “Light field quality assessment datasets - what is missing?” by Maria Martini of Kingston University London

#210 “A new light field quality assessment dataset” by Maria Martini of Kingston University London

#211 “Status and summary of IEEE recommended practice for the quality assessment of light field imaging” by Maria Martini of Kingston University London

QAH

#221 “New insights on Affinity therapy for people with ASD: an eye-tracking study” by Julie Fournier of INSA Rennes

ASD: autism spectrum disorder. Affinity therapy: key opening to the world / language / etc.

Eye-tracking experiment on natural images and images of interest (viewer dependent).

Question from Lucie: saliency maps? → Ongoing work! Future challenges.

#222 “Evaluation on the usability of deep learning based denoising models for low dose CT simulation” by Lumi Xia of INSA Rennes

(NB) Recording requested to be removed

Compromise between radiation dose (risk) and image quality.

Adapt DL-based denoising models for low dose CT simulation: ResNet and U-Net.

Metrics to assess the similarity of noise level (histogram correlation and mutual information).

Comment from Ioannis about metrics: maybe try Kullback-Leibler divergence.

Question from Lucie: subjective tests? Future work.

#237 “Deep-based QA of medical images through domain adaptation” by Mohamed Amine Kerkouri of University of Orleans

Based on a paper presented at ICIP 2022: <https://arxiv.org/abs/2210.10533>.

Goal: to analyze the quality of scan, to understand immediately if it will need to be re-taken, instead of long wait for review by a medical professional.

Datasets: **MD72** (medical images) and **TID13** (general images).

Proposed model: SAQM (self attention quality metric).

Question from Jorge: assumption of quality only dimensional? Also, there might be differences between normal and pathological images… and when images are ambiguous… → Amine just checking bottom-up features / visual quality. Need for more datasets!!!

Question from Margaret: where can we find the **MD72 dataset**? (very good question!)  
→ MD72 dataset: <https://drive1.demo.renater.fr/index.php/s/jFR5J6j3fWFAet4>

Paper associated to this dataset: <https://pubmed.ncbi.nlm.nih.gov/30129565/>

#233 “Cognition Inspired Diagnostic Image Quality Models” by Jorge Caviedes of ASU

Based on a paper presented at ICIP 2022.

More complex story for medical image quality: interpretability, adequacy, and visual quality!

Distinguish between:

* **Interpretability** (e.g., medical professional is confident in making a diagnosis)
* **Adequacy** (e.g., capture technique shows the right area for assessment, person held still as instructed during imaging)
* **Visual quality** (e.g., MOS)

All of these are content dependent. Prototype web-based subjective test interface, with these three questions. Also asks subject to explain any ratings that are less than perfect.

Objective diagnostic image quality (DIQ) metric based on the 3 dimensions.

Questions from Margaret: make interface available for subjective tests and dataset?  
→ Looking for a way to do this. Students interested but no funding.

#223 Discussion on the future of QAH

“Objective quality assessment of medical images and videos: Review and challenges”: <https://arxiv.org/abs/2212.07396> (submitted to *Multimedia Tools and Applications*)

Opportunity to conduct medical perception research shared by Elizabeth Krupinski: during the European Congress of Radiology (ECR 2023), which will be held in Vienna in March 2023. Please note any questions should be directed to Jeremy Wolfe ([jwolfe@bwh.harvard.edu](mailto:jwolfe@bwh.harvard.edu)) and Todd Horowitz ([todd.horowitz@nih.gov](mailto:todd.horowitz@nih.gov)). [More information here](https://drive.google.com/file/d/13sVB-v9-bz4TH6TRTwkge8VVRB9JxJ5Q/view?usp=sharing).

Basically: we want “fresh blood” (new people, new ideas) to join us!

For more info: <https://vqeg.org/projects/quality-assessment-for-health-applications-qah/>

Contact [qah@vqeg.org](mailto:qah@vqeg.org) and fill in [this Google Doc](https://docs.google.com/spreadsheets/d/1ckjc67QKOc9NyKcRZRTBLh1SU4fuGyxFUCdEaWtBOdg/edit#gid=0)

We plan a discussion January: <https://doodle.com/meeting/participate/id/dGv4MyKb>

**Minutes are approved to this point.**

The meeting recordings will be made available in the shared google drive folder: [https://drive.google.com/drive/folders/1vEDujX\_THFEW6fLBtF1IXJdV2AKAj6Ro?usp=sharing](https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdrive.google.com%2Fdrive%2Ffolders%2F1vEDujX_THFEW6fLBtF1IXJdV2AKAj6Ro%3Fusp%3Dsharing&data=05%7C01%7Cmpinson%40ntia.gov%7Ca07ad3a63fa540671ccd08dadbacc2db%7Cd6cff1bd67dd4ce8945dd07dc775672f%7C0%7C1%7C638063832332476031%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=E5L0%2BS4sjSjg83biplEcGafLDqqvypzkS2L9ve2YtKM%3D&reserved=0)

Day 4 (Thursday, December 15)

JEG-Hybrid

#218 Updates on JEG-Hybrid activities

Presented by Enrico Masala

Currently working on research problems rather than algorithms/models with immediate applicability. For more details, see the next two presentations.

New JEG-Hybrid website (also linked from main VQEG website) <https://vqegjeg.github.io/jeg-hybrid/>

Includes: list of activities of interest, freely downloadable publications, resources

#219 On the Training of AIOs for a wider range of Applications

Presented by Lohic Fotio Tiotsop

Presented results of DNNs trained to model single observers (so-called Artificial Intelligence Observers - AIOs): how to train the network using a mix of automatically-annotated large datasets and transfer learning using smaller subjectively-annotated datasets.   
Presented new results that included not only JPEG distortions but also noise, blur, JP2K distortions.   
Showed better robustness to image perturbation (small Gaussian noise, conversion to grayscale). Also, showed that AIOs are able to retain bias and inconsistency when trained on a subjectively annotated dataset where bias and inconsistency have been artificially added, and then tested on a different, large scale dataset.

Ioannis: particularly likes the reverse mapping between the quality parameters and the MOS scores

Ioannis: also suggests looking at using intermediate values between the five point scale also during DNN training (i.e. considering more than 5 intervals for the distortion control parameter), and compare the results to the one using only the five point scale. Also, use networks that predict a continuous value and not only probability for each one of the 5 quality points on the scale and then derive the average, and compare results.

#220 Guidelines to Train, Validate and Publish DNN-based Objective Measures (Ongoing Work)

Presented by Lohic Fotio Tiotsop

Aim to write a journal paper proposing a template to facilitate reporting results about DNN-based quality assessment measures. Hopefully evolving it into an international standard. Presented how even small details matter, e.g., how you transform a full image into the input patch required by the DNN, or how much results may change even with small input perturbation (such as small Gaussian blur).  
Work is ongoing and open to anybody interested, please join JEG calls to be involved: <https://vqegjeg.github.io/jeg-hybrid/meetings>

Ioannis: it is important to do sensitivity tests with objective metrics to ensure stability when there are small perturbations: metrics should behave coherently. This does not make sense for subjective testing (because of costs etc.) but for objective metrics should definitely be done, there is no reason why this should not be done.

International Recommendation: Margaret suggests to involve ITU-T SG12 as early as possible in the process. It is much easier to reach a recommendation when they are involved in the process rather than present a full “finished” work and then ask for discussion and acceptance. Pablo is a member of SG12 and willing to help/be involved, Kjell suggested that Jörgen Gustafsson (Ericsson) might also be interested and willing to help.

Ioannis: suggested to include in the template simple but effective complexity metrics (see also PCS 2022), in particular memory footprint in MB, and a standardized proxy of complexity, independent of network architecture, i.e., the number of operations MUL-ACC (MAC) per pixel which are needed to run the model.

IMG

#228 UX and QoE aspects of remote control operations using a laboratory platform (work in progress)

By Shirin Rafiei of RISE Research Institutes of Sweden

Application of immersive media to remote control vehicles, the goal is to characterize the system characteristics necessary to ensure sufficient QoE to complete the task. Example given is logging, but the results could be applied to other areas.

Margaret: Question on the fisheye lenses, with strong distortion for the small lab model used for the experiment? Shirin and Kjell: Does not hinder navigation. Real-world users want fisheye lenses as well, to better understand the surrounding area.

Antonio Servetti and Enrico: consider that some cameras might introduce 50-60 ms delay because of their inner workings (pipelining of internal operations, USB isochronous transfers, …). In the past we had our best results using industrial cameras (that output low-latency, not compressed content).

#224 Behavioural Analysis in a 6-DoF VR System: Influence of Content, Quality and User Disposition

Understanding user behaviors when interacting with VR systems and able to move with 6-DoF. Prior work limited to moving only the head (3-DoF).

#226 Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds

<https://github.com/cwi-dis/cwipc>

#229 “IMG next joint effort” by Jesús Gutiérrez, Pablo Pérez of Universidad Politécnica de Madrid, Nokia

Plan to study improved protocols for interactive subjective testing. **Welcome new participation**. This will start holding monthly calls. Each lab designs their own experiment. Will rely on the VQEG SAM group to recommend statistical evaluation methods. Monthly calls will be announced on the IMG reflector.

**Anyone who is interested:** send an email to the IMG reflector, or directly contact the IMG co-chairs.

Schedule:

* Planning Dec. 2022 to May 2023
* Execution May 2023 to Dec 2023
* Evaluation Dec. 2023 to May 2024.

Basically, the goal is to move from one phase to the next at Spring and Fall VQEG meetings.

#231 “PointPCA: Point Cloud Objective Quality Assessment Using PCA-Based Descriptors” by Irene Viola of CWI

<https://arxiv.org/abs/2111.12663>

#232 “Comparing ACR, SSDQE, and SSCQE in long duration 360-degree videos” by Marta Orduna of Universidad Politécnica de Madrid

Interested in interactivity, naturalness, and reproducibility for three assessment methods: ACR (short duration), SSCQE (continuous, long content), and SSDQE (long content).

Results will be used to submit a proposal to extend ITU-T Rec. P.919 for long sequences.

AVHD

#212 “Converting video QA metrics -  A simple relationship between SSIM and PSNR for DCT-based compressed images and video: SSIM as content-aware PSNR” by Maria Martini of Kingston University London

Establish and demonstrate the relationship between PSNR and SSIM. This is a simple relationship as they are linked by a simple content related no reference factor.

Ioannis: the assumptions are definitely valid and further validate the use of local variance to do adaptive-QP (AQP) encoding in x264/x265 and other popular video SW encoders. It would be interesting to check the accuracy of the approximation by using source local variance (instead of distorted image variance) and also to incorporate a global correction factor for the lower variance of compressed images (especially in low QP/bitrates)

#215 “Deep dive into Video Codec Profiling with VQA complexities and resolutions” by Urvashi Pal of Akamai

Complexities that a company faces to do viewing in the lab, especially when there are multiple codecs and codec vendors and codec profiles, with different requirements.

**These minutes have been approved to this point.**

Day 5 (Friday, December 16)

SAM

#238 Collective Just Noticeable Difference Assessment for Compressed Video with Flicker Test and QUEST+

Presented by Mohsen Jenadeleh

Four JND datasets identified.

Propose method to improve sensitivity and precision.

Ioannis raised that such test is very artificial and the obtained results can overestimate the needed quality.

Kjell pointed to publication with flicker test run by VQEG some time ago: Perspectives on the definition of visually lossless quality for mobile and large format displays, Journal of Electronic Imaging, DOI: 10.1117/1.JEI.27.5.053035 with link <http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Ari%3Adiva-35314>

Dawid asked about more details of the crowdsourcing platforms.

Lucjan asked about individual testers differences if they were calculated. They were not.

Interlude for Administrative discussion

Topics: (1) creation of new group, (2) whether to close existing groups PsyPhyQA and AVHD, and (3) proposed liaison statement

**Decision was reached: Agreement** to create a new VQEG group: **Emerging Technologies Group (ETG)**

**Chair and Vice-Chair:** Nabajeet Barman and Saman Zadtootaghaj

Saman may be replaced by Marcos V. Conde; we want him to be present at a meeting before make this change.

Text for VQEG website is here: <https://docs.google.com/document/d/164EQ02T3KtPmcZQf3_pR1AXIz4Vryd0N--jiPKilku0/edit>

Recordings and YouTube channel for VQEG. Nabajeet has access. Femi listed as responsible for this.

Creation of VQEG slack channel? To supplement email reflectors.

**Agreement to close PsyPhyQA** until interest resumes in this effort.

**Send email to AVHD co-chairs,** ask for their input on whether to continue or close this group or change scope. Will likely wait until the next VQEG meeting to make a decision. If AVHD is closed, we would need another place where FR metrics and JND assessment are discussed.

Implementer's Guide for Video Quality Metrics (IGVQM) closely related to new direction of JEG-Hybrid work. Propose to move IGVQM into JEG-Hybrid will be discussed by JEG-Hybrid, and decisions reported to VQEG Board.

Discuss policy for incoming and outgoing liaisons between VQEG and AOMedia. Confirmed that these liaisons will be publicly available, as per VQEG policy 3.3.

#204 “Updating and merging P.913, P.911, and P.910” led by Margaret Pinson of NTIA/ITS

Baseline document for discussion is VQEG\_SAM\_2022\_T-REC-P.913\_draft\_changes\_accepted.docx and VQEG\_SAM\_2022\_T-REC-P.913\_draft\_changes\_marked.docx

Modified document with our copies will appear here:

VQEG\_SAM\_2022\_T-REC-P.913\_draft\_changes\_marked\_ver2.docx

See also “draft liaison statement to ITU-T SG12.docx”

The updated liaison and document sent to ITU-T SG12 will be available in the meeting files.

**The minutes have been approved to this point.**

**The VQEG meeting officially closes.**