VQEG Meeting Minutes

14 December 2020

Project Summaries

File: VQEG\_ADMIN\_2020\_148\_Status\_update

Psycho-Physiological Quality Assessment (PsyPhyQA)

“Exploring Neural and Peripheral Physiological Correlates of Simulator Sickness”

by JP Tauscher

File: VQEG\_PsyPhyQA\_2020\_135

“Visual fidelity of tone mapping operators from gaze data using HMM”

By Waqas Ellahi/ toinon Vigier/Patrick Le Callet

File: VQEG\_IMG\_2020\_140

Computer Generated Imagery (CGI)

“Updates on gaming standardization activities and deep learning models for gaming quality prediction (DEMI, NIMA)”

By Saman Zadtootaghaj

File: VQEG\_CGI\_2020\_110

Qn from Ioannis (FB): Is there any consideration of any other non-pixel related factors such as frame drop, audio quality, etc.

    Qn from Lucjan (AGH):

    Remark from Sahid: Some people prefer blockiness, while some prefer blurriness - at least this is     the observation from subjective tests done on natural content.

    Follow up remark by Ioannis: VMAF does try to take-into account the trade-off between blockiness and blurriness - was there a particular distribution observed?

    No fixed study done - but implicit observation is that content might have an effect on preference but in general blurriness might be slightly preferred over blockiness.

“A subjective study of multi-dimensional aesthetic assessment for mobile game images”

By Junle Wang, Suiyi Ling

File: VQEG\_CGI\_2020\_116

Presented by Suiyi Ling:

Aesthetic quality plays an important role in the gaming design process.

Four dimensions are considered for quality assessment of mobile gaming quality:

* Fineness
* Colourfulness
* Color harmony
* Overall aesthetic

Experiment: ACR  5-point rating scale from

Correlation of overall quality to other dimensions: colorfulness turned out to be the most important dimension.

Correlation of Confidence Intervals between the dimensions revealed a very similar result (high correlation of overall quality with colorfulness)

    Qn from Ioannis: How was the correlation between the two CIs calculated?

        CI values were calculated for different measurements and then the corr values are       calculated between the CI values. (It seems the width of the CIs are taken into account.) One can say that the correlation between CIs widths expresses the relationship between the uncertainty of two factors.

Qn from Saman: was there any compression effect taken into account?

    Not for this work, ongoing

        The dataset is available (link in the paper)

    Suggestion from Mikolaj: It will be good to explore the intercultural correlation between ratings/preferences.

    Qn from Ali (Uni Nantes): Any categorical differences observed between the categories considered and the dimensions considered (along with the overall quality).

    Not but looks like an interesting option to explore further.

“Quality assessment of gaming videos compressed via AV1”

By Maria Martini

File: VQEG\_CGI\_2020\_129

Comparison of AV1, HEVC and H.264 on gaming video dataset .

AV1 results in the best quality for most bitrate and content.

The study was conducted with 15 participants.

VMAF performs the best among the metrics tested.

Qn from Shahid: you mentioned gaming content is different compared to natural content which affects the codec performance for different content.

Maria: We observed this also in previous work (we presented the relevant observations  in QoMEX2018).

Qn from Shahid: SSIM performs better for HEVC compared to other codecs. Is there any reason for that?

Maria: It might be because HEVC compression preserves better the structural similarity, which is measured by SSIM

Comment from Ioannis: PSNR works better when you have more high quality content, perhaps it performs better with AV1 since the quality is higher.

Comment from Ioannis: the VMAF is trained on the H.264 content and we can see that it performs well with the AV1 codec as well, which is a relief.

Pablo: This is not the first time showing the VMAF works very well with gaming content.

Maria: We have seen that also in our previous work. This might be due in part to the fact that VMAF used some CGI content for training.

Zhi li: We used indeed some animation, although not gaming content.

Nabajeet: Also, VMAF is for video, while the other metrics were developed for images.

Qn: please provide the link for Gaming Quality Datasets used in this paper.

Nabajeet: We provide the link of all datasets in the minutes:

The dataset used in the paper presented (above) is the GamingVideoSET available here: <https://kingston.box.com/v/GamingVideoSET>

* Relevant dataset paper: <https://ieeexplore.ieee.org/abstract/document/8463362>

Some other additional Gaming QoE related datasets (briefly presented by the chair, Saman - see presentation VQEG\_CGI\_2020\_110)

KUGVD: [https://kingston.box.com/v/KUGV](https://kingston.box.com/v/KUGVD)D

* Paper: <https://ieeexplore.ieee.org/iel7/6287639/8600701/08727887.pdf>

CGVDS: <https://github.com/stootaghaj/Cloud-Gaming-Video-Dataset>

* Paper: <https://dl.acm.org/doi/abs/10.1145/3339825.3391872>

GISET: <https://github.com/stootaghaj/GISET>

* https://link.springer.com/article/10.1007/s11042-020-09144-6

Immersive Media Group (IMG)

“AVT Team Update on Immersive Media topics”

By Stephan Fremerey (on behalf of AVT)

File: VQEG\_IMG\_2020\_117

Qn from Ioannis: Motivation to evaluate differences among multiple encoding phases.

“Perceptual Characterization of 3D Graphical Contents based on Visual Attention patterns ”

By Mona Abid, Matthieu Perreira Da Silva, Patrick Le Callet

File: VQEG\_IMG\_2020\_146

Paper link: <https://dl.acm.org/doi/abs/10.1145/3423328.3423498>

Suggestion of  a way to **perceptually characterize 3D Graphical Contents (3DGC)** based on Visual Attention.

Based on collected human gaze data (obtained from an eye-tracking experiment on rendered 3D objects), we revised the **Visual Attention Complexity (VAC)** measure, initially suggested in 2D imaging, and we adapted it to 3D objects.

● Subjective study:

■ Dataset + Eye-tracking data: <https://gitlab.univ-nantes.fr/E18E421U/3dgc-dataset>

■ Adapt VAC measure for 3DGC : **VAC-3DGC**

We also provided an objective predictor that highly mimics the experimental Attentional Complexity information.

● Objective predictor

■  View-wise analysis & 3D Object-wise analysis

■ This predictor can be useful in Quality of Experience (QoE) studies:

* to balance content selection when creating a 3D dataset
* and to optimize the user’s QoE during interactive visualization (i.e. quantifying the attentional complexity of each viewpoint of a 3D object)
* More recently, integrated in a Perceptual Quality Metric for 3D Graphical Contents (Future Publication )

“IMG Work Plan - What's next?”

Jesús Gutiérrez / Pablo Pérez

File: VQEG\_IMG\_2020\_139

Discussion to be continued in the next IMG audio-call.

15 December 2020

IRG-AVQA

chaired by Chulhee Lee

Overview

By Chulhee Lee, Jens Berger, Kazuhisa Yamagishi, Alexander Raake, Ioannis Katsavounidis

File: [IRG-AVQA 2020-12.pptx](ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Stockholm_Dec20/IRG-AVQA%202020-12.pptx)

<https://www.itu.int/md/R19-WP6C-C-0051/en>—ITU-R WP6C  Contribution  51; HDR-related proposal (currently in the early stage).

Adolph, Martin:

Perhaps just to add to Q13 - ITU-T has recently approved new Recommendation ITU-T P.919: Subjective test methodologies for 360º video on head-mounted displays, in collaboration with VQEG IMG

P.NATS Phase 3 is starting—long-term integration.

Babak Naderi:

For Q7 there is a new work item P.CrowdV for "Subjective evaluation of video and audiovisual quality with the crowdsourcing approach". We are looking for contributions here. <https://www.itu.int/ITU-T/workprog/wp_item.aspx?isn=16541>

Ioannis’ proposal for introducing other-than-PSNR metrics to the video compression community.

Everyone is invited to contribute to this effort.

Potential database that could be used: <https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-1>

**Let’s set up an email reflector; dedicated group and start the discussion ASAP**. Interested parties should contact Ioannis directly through email.

**Interested labs:**

* NTIA (Margaret Pinson),
* Rhode & Schwarz (RS) SwissQual AG,
* University of Nantes (Patrick Le Callet),
* Yonsei University (Chulhee Lee),
* RISE Research Institutes of Sweden AB (Kjell Brunnstrom),
* AGH University of Science and Technology (Mikolaj, Lucjan, Jakub, Dawid),
* University of the West of Scotland (UWS) (Naeem Ramzan),
* Facebook (Ioannis),
* TU Ilmenau (AVT) (Steve Goering),
* Universidad Politécnica de Madrid (UPM) (Jesus Gutierrez),
* Wuhan University (Zhenzhong CHEN, Yaosi Hu),
* Tencent Media Lab (Joel Jung)

Decision Reached: New VQEG Group established:   
Implementers Guide for VQM (IGVQM)   
Chair: Ioannis

Outputs: guides for video quality metrics

From the chat thread:

* [6:22 AM] steve.goering
* this db couldalso be used<https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-1>
* [6:23 AM] steve.goering
* 4 tests are there included with resolutions ranging up to 4K
* [6:23 AM] steve.goering
* (or better UHD-1)
* [6:40 AM] Naeem Ramzan (Guest)
* We are interested
* [6:42 AM] Kjell Brunnström
* Put us on the list too
* [6:43 AM] steve.goering
* TU Ilmenau (AVT) would also be interested
* [6:44 AM] jesus.gutierrez@upm.es (Guest)
* we are interested too: UPM
* [6:44 AM] Zhenzhong CHEN (来宾) (Guest)
* Ioannis (Guest) we are interested in participating (i am not able to use the mic)
* [6:44 AM] Zhenzhong CHEN (来宾) (Guest)
* we （Wuhan University）are interested in participating (i am not able to use the mic)
* [6:46 AM] Zhenzhong CHEN (来宾) (Guest)
* Ioannis (Guest) please put Yaosi Hu to represent for Wuhan University (WHU)
* Ioannis (Guest)thanks a lot
* [6:47 AM] Mikołaj Leszczuk (Guest)
* Ioannis, it seems that actually AGH contribution could be a bit wider. Details TBA. i need to discuss this with Lucjan and Jakub.
* [6:53 AM] Joel Jung (Guest)
* Ioannis, you can also add my name (Joel Jung), on behalf of Tencent Media Lab

VQEG will draft the guide and submit to the ITU

Q19 Interim Meeting

chaired by Chulhee Lee

Alternating Project Solver

By: Zhi Li (Netflix)

File: [VQEG\_Q19\_2020\_136 Improvements on Subjective Experiment Data Analysis Process An Update (Q19 Interim Meeting, VQEG Fall 2020).pdf](ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Stockholm_Dec20/VQEG_Q19_2020_136%20Improvements%20on%20Subjective%20Experiment%20Data%20Analysis%20Process%20An%20Update%20(Q19%20Interim%20Meeting,%20VQEG%20Fall%202020).pdf)

Lohic: can this model better predict subject responses than the MOS? (to be continued offline between Zhi and Lohic).

Statistical Analysis Methods (SAM)

“suJSON: A Uniform JSON-based Subjective Data Format”

By Jakub Nawała

File: VQEG\_SAM\_2020\_106\_suJSON.pdf

JN: I encourage all academics to invite their students to contribute to this project.

“Describing Subjective Experiment Consistency by p-Value P–P Plot”

By Jakub Nawała

File: VQEG\_SAM\_2020\_105\_Experiment\_Consistency\_ p-Value\_P-P\_Plot.pdf

Extend this to detect outlying subjects by performing the analysis per-subject (i.e., generating per-subject score distributions).

<https://dl.acm.org/doi/10.1145/3394171.3413749>—the paper in ACM Digital Library (along with the video recording of the presentation)

<https://arxiv.org/abs/2009.13372>—the paper on arXiv

Propose a single measure assessing the consistency of the whole experiment.

This can be applied to other models (other than the Generalized Score Distribution). For example, one can apply the methodology with the model proposed by Zhi, Lucjan and Ioannis.

Patrick Le Callet

It may not be the best idea to say that some score distributions are typical or atypical. Different use cases may yield different score distributions. The method should thus be adapted to a given use case.

“How to Define the Quality of a Single Sequence?”

By Lucjan Janowski

File: VQEG\_SAM\_2020\_114

<https://create.kahoot.it/share/vqeg/087a11c8-4e1e-4cc0-9f26-56b2d82fae54>

Kamil Koniuch: Can quality exist without a person?

Pablo Perez: There is no true quality of a single stimulus.

Maria Martini:

We have [a related paper in Transactions on Multimedia](https://eprints.kingston.ac.uk/id/eprint/46518/1/Martini-M-46518-1.pdf). (we called it PMOS - population MOS - and aimed at estimating it in the paper)

Perceived quality vs objective quality?

Alex Raake:

Try Sensory Evaluation techniques (take a look at D. Keller’s QoMEX 2019 paper). Human factors cannot be probably completely ruled out of the equation. Quality as an implicitly human concept. Internal points of reference matter.

Lukas Krasula

Quality seems to be relative (i.e., no absolute quality?).

Lucjan Janowski

If we cannot agree what quality means then what do we mean by “objective quality” when building objective metrics?

In general, quality depends very much on the context.

“Mean Opinion Score and Ranked based Statistics”

By Babak Naderi

File: VQEG\_SAM\_2020\_125

See <https://github.com/babaknaderi/MOS-transformation>

For code implementing this technique

Patrick Le Callet

By transforming the results artificially we detach ourselves from the theory standing behind statistical significance tests. In other words, we are at risk of going beyond the range of applicability of statistical methods.

Immersive Media Group (IMG)

“no reference quality evaluation of light field content based on structural representation of the epipolar plane image”

By Ali Ak and Patrick Le Callet

File: VQEG\_IMG\_2020\_141

When the source code is available it will be posted on the VQEG NORM/IMG’s web site.

Wednesday, December 16

CACoViA Session

“Methods for Objective Video Quality Assessment in Recognition Tasks”

By: Mikołaj Leszczuk, Jakub Nawała, Lucjan Janowski, Atanas Boev

File: VQEG\_QACOVIA\_2020\_104 (will be shared in Teams meeting’s chat)

Ioannis Katsavounidis

Most of the indicators you use operate only on the luma channel. Please remember to mention this in the subsequent publication.

Margaret Pinson

The camera noise model is not the same as Gaussian noise. Indicators that work well for Gaussian noise do not work well for real camera noise.

Atanas Boev

Does anyone know about real camera noise models?

Alessandro Foi from *Noiseless Imaging* was working on a camera noise model for research

I. Zabalta, M. Camara, C. Diaz, T. Canham, N. Garcia, M. Bertalmio, “Retinal noise emulation: a novel artistic tool for cinema that also improves compression efficiency,” IEEE Access, vol. 8, pp. 67263-67276, Apr. 2020.

K Kawa, M Leszczuk, A Boev, “Survey on the State-Of-The-Art Methods for Objective Video Quality Assessment in Recognition Tasks”, International Conference on Multimedia Communications, Services and Security …, 2020, <https://link.springer.com/chapter/10.1007/978-3-030-59000-0_25>

Databases (potentially) resulting from the grand challenge organised by NTIA/ITS may be useful here. These databases will be available no earlier than June 2021.

5GKPI

Pablo gives an Introduction to 5G KPI

5G enables new video technologies, triggering new video quality questions.

Presentations

“Tele-operated Driving Use Cases”

By: Yunpeng Zang

File: VQEG\_5GKPI\_2020\_120\_5GAA\_ToD

Ruiz Alonso

Excellent presentation from Yunpeng presenting ToD from 5GAA ! Hope to collaborate soon from 5G-MOBIX with low latency video tests..

Ioannis Katsavounidis

Do you really need 36 Mbit/s in the uplink (vehicle to remote driver) to reliably stream high quality video? SoTA codecs should do the job well with less throughput available.

Margaret Pinson

Can you please share a few representative videos so that VQEG could see in practice what use-cases are relevant.

A dataset of representative videos would help VQEG explore quality issues.

Lucjan Janowski

Can you actually test the influence of video quality without asking people to remotely drive a real car?

**Answer:** It depends on the use case. Sometimes a toy car is sufficient (e.g., to test communication infrastructure), sometimes you need a real car (e.g., if you want to learn more about safety concerns). In general, however, no simulated driving or no real remote driving does not seem to be applicable.

Link to an experiment using a toy car: <https://www.youtube.com/watch?v=6Sfg-JZYh9E&feature=emb_title>

“5G for Content Production”

By: Paola Sunna, Norbert Werner

File: VQEG\_5GKPI\_2020\_121

Two technical reports (TRs) on 5G created and published recently by the EBU:

* <https://tech.ebu.ch/publications/tr054>
* <https://tech.ebu.ch/publications/tr056>

Developing objective models predicting subjectively perceived quality for both audio and video are still very relevant.

Ioannis Katsavounidis

Is it really necessary to focus solely on CBR? It’s understandable for DVB-T applications, but for packet networks it does not seem to be necessary.

Ian Wagdin

Variable BitRate (VBR) may not be that relevant since in most of the professional cases the codec is operating with maximum possible bitrate. In this case: VBR = Constant BitRate (CBR).

3GPP has authored documents that may be interesting here:

<https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3648>

<https://tech.ebu.ch/publications/tr056>

“Quantifying the value of 5G and edge cloud on QoE for AR/VR”

By: Bill Krogfoss

VQEG\_5GKPI\_2020\_122

Link to paper:

<https://ieeexplore.ieee.org/document/9123090>

Quality Assessment for Health Applications

“Introduction”

By: Lu Zhang

File: VQEG\_QAH\_2020\_133

“On the subjective assessment of the perceived quality of medical images and videos”

By: Lucie Lévêque

File: VQEG\_QAH\_2020\_107

QAH will make list of datasets available for research

French and China radiologists had different bias on MOS. A topic for future discussion perhaps: different communities have very different standards for acceptability.

For health applications, must make sure impairments match actual use cases (realistic). Real-time environment mentioned here makes sense, because the use case is time-critical response for live intervention.

“On the suitability of VMAF for quality assessment of medical videos: Medical ultrasound & wireless capsule endoscopy”

By: Maria Martini

File: VQEG\_QAH\_2020\_128

Link to paper:

<https://www.sciencedirect.com/science/article/pii/S0010482519302604?casa_token=Uio9DSYtBDYAAAAA:mLHBNKjsdRgu5Ym55dJ_17WuwHOAQQ1hvSTmwt_up-r19fa6nHL2vvOrWkdvTPPC6SctyFVl>

“Cognition inspired diagnostic image quality models”

By: Jorge Caviedes

File: VEQG\_QAH\_2020\_131

AVHD Session

“Hardware acceleration of video quality metrics”

By: Cindy Chen et al.

File: VQEG\_AVHD\_2020\_118

SPIE paper: <https://spie.org/Publications/Proceedings/Paper/10.1117/12.2569302>

“Efficient Measurement of Quality at Scale in Facebook Video Ecosystem”

By: Haixiong Wang

File: VQEG\_AVHD\_2020\_126

SPIE paper link: [https://spie.org/Publications/Proceedings/Paper/10.1117/12.2569920](https://spie.org/Publications/Proceedings/Paper/10.1117/12.2569920?SSO=1)

[Ioannis, Haixiong] please contact us directly for further questions/comments

Wednesday minutes approved.

Thursday 17 December

JEG-Hybrid

“Comparing commercial and open source VQMs for HD constant bitrate videos”

By: Florence Agboma and Lohic Fotio Tiotsop

File: VQEG\_JEGHybrid\_2020\_109

Interesting results on subject screening, comparison of methods / decisions.

When all of the metrics agree on the quality rating of a PVS, their predictions seem to closely match subjective testing MOS.

Exploring implications when multiple metrics disagree on the quality of a sequence. This seems to be related to how the video is encoded.

Ioannis suggested to remove the black bands before computing the objective measures, as differences might be “washed out” since quality for those bands is almost perfect.

Kjell: NTIA VQM could also be included in the comparison

“Artificial intelligence-based observers for media quality assessment”

By: Lohic Fotio Tiotsop

File: VQEG\_JEGHybrid\_2020\_108

Interesting approach for shallow NN case.

Adding VMAF to the comparison would be beneficial. Also, mixing full-reference and no-reference measures do not help.

Ioannis: skeptical about being able to train DNNs (even just transfer learning) with few (hundreds) of ground truth data points for single observers (information theory considerations)

“Comparing full-reference video quality metrics using cluster analysis”

By: Dariusz Grabowski    AGH University of Science and Technology

File: VQEG\_JEGHybrid\_2020\_137

AVHD Session

“AVHD-AS / P.NATS Phase 2 Processing Chain”

By: Werner Robitza, David Lindero

File: VQEG\_AVHD\_2020\_103\_pnats2avhd-processing-chain

Software available on GitHub: <https://github.com/pnats2avhd/processing-chain>

This is the processing chain used to generate sequences for the P.NATS Phase 2 / AHVD-AS project from ITU-T SG12 and VQEG.

A sample database is available at:<https://github.com/pnats2avhd/example-databases>

Zhi Li: Can you generate encoded videos on-the-fly?

**Ans.** No. Everything has to be generated prior to a subjective experiment.

The presentation also points to the software package inserting stalling events and initial buffering: <https://github.com/slhck/bufferer>.

Video playback was implemented using Blackmagic Design playback/capture cards (for the TV setup) + OPICOM’s playback software (available through [the VQEG Tools web page](https://vqeg.github.io/software-tools/tools/)).

“P.NATS Phase 2: ITU-T Recommendation P.1204 (P.1204.3, P.1204.4 & P.1204.5) - Multi-Model Standard for Bitstream-, Pixel-Based and Hybrid Video Quality Assessment of UHD/4K”

By: Alexander Raake, Shahid Satti, Silvio Borer, Jörgen Gustafsson, Rakesh Rao Ramachandra Rao

File: VQEG\_AVHD\_2020\_127

Repo with the software standing behind P.1204.3: <https://github.com/Telecommunication-Telemedia-Assessment/bitstream_mode3_p1204_3>

A link to additional data sets used to test P.1204.x models: <https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-1>

AGH is interested in joining the “MOSQUITO”(?) ITU work item.

There were no stallings or initial delays in all the PVSs.

SAM Session

”On VMAF's property in the presence of image enhancement operations”

By: Zhi Li

File: VEQG\_SAM\_2020\_124

Ioannis

Should we include image enhancement in subjective testing when we do research on subjectively perceived quality?

**Ans.** That’s difficult to resolve. Let’s perform relevant subjective experiments and see how that works.

The largest Image Aesthetic Assessment (IAA) data set to date:

N. Murray, L. Marchesotti and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 2408-2415, doi: 10.1109/CVPR.2012.6247954.

Related paper:

L. Krasula, M. Narwaria, K. Fliegel, P. Le Callet “Influence of HDR reference on observers preference in tone-mapped images evaluation,” 7th International Workshop on Quality of Multimedia Experience (QoMEX), 2015

Pablo

Content producers don’t want “transmission guys” to enhance their content. This is at risk of distorting the initial artistic intent.

Maria Martini

Keep in mind that if it comes to medical imaging you never deal with artistic intent. However, some image enhancement may be beneficial for diagnosis. Nevertheless, users must be informed that some enhancements have been applied.

“Confidence intervals for subjective tests & objective metrics”

By: Margaret Pinson

File: [VQEG\_SAM\_2020\_112](ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Stockholm_Dec20/VQEG_SAM_2020_112_Confidence_intervals_for_metrics.pptx)

Margaret H. Pinson, "[Confidence Intervals for Subjective Tests and Objective Metrics That Assess Image, Video, Speech, or Audiovisual Quality](https://www.its.bldrdoc.gov/publications/3253.aspx)," NTIA Technical Report TR-21-550, October 020.

<https://www.its.bldrdoc.gov/publications/details.aspx?pub=3253>

Code: GitHub, <https://github.com/search?q=normetricframework>

See function ci\_calc.m or ci\_calc.py

“A graphical probabilistic model to recover the ground truth and annotator's behavior”

By: Jing Li，Suiyi Ling, Patrick le Callet, Junle Wang, Zhi Li

File: VQEG\_SAM\_2020\_119

Accepted to ACM Transactions on Multimedia.

AGH proposed to join forces with Jing, Suiyi, Patrick, Junle and Zhi.

5GKPI

“5G and QoE for remote controlled use cases”

By: Kjell Brunnström

File: VQEG\_5GKPI\_2020\_123

Friday 18 December

NORM

Ioannis (Facebook) on new codec development

50 to 60 sequences, 4 to 10 seconds, selected to train new codec

Diversity in pixel domain (dark, light, detailed, fast motion,...) corner cases of video complexity will help development of next standard; difficult content where prior codec fails.

Artistic intent and user generated content excluded. Very popular (undeniably). Previously compressed content, uploaded to website, debatable whether subsequent encoding would increase or decrease quality. Unsure how to improve.

Need a good NR metric (perfect not needed), because PSNR is not much more accurate than NR metric.

* UGC content is very important
* Problem is on the metric side

“Adventures in Implementing SI and TI”

By: Cosmin Stejerean

File: VQEG\_NORM\_2020\_111

P.910 has ambiguities, this results in different values computed for the same sequence by different software.

Werner Robitza and Steve Goering authored SI/TI software, contributed to discussions, different public implementations are:

* <https://github.com/slhck/siti> (Python library with support for YUV and limited/full range)
* <https://github.com/Telecommunication-Telemedia-Assessment/SITI/> (OpenCV/C++ library performing RGB conversion, no support for limited/full range)
* <https://github.com/Telecommunication-Telemedia-Assessment/SITI/tree/master/python> (Python version of the above)

Cosmin made a proposition to generate (and maybe standardise) several test vectors. These would serve as reference values to which one could compare to when implementing their own TI/SI calculating solution.

And also to modify P.910 to clarify the ambiguous details. Suggestion from Steve to discuss with TU Ilmenau/AVT Group (ITU members) to bring in proposed changes.

Create a good reference implementation, fixed point, “this is exactly how you do it” avoid floating point with different precision in different implementations.

What do we do about HDR? The question is not just 10 bit, but what type of luma used for HDR?

General question (Werner) about usefulness of SI/TI raised by some video engineers he talked with, why not just use CRF and look at the file size, or some function of CRF, SI, and TI? CRF would be more meaningful for gauging complexity specifically for encoding tasks.

Discussion: Audio call in January

* Werner
* Zhi
* Cosmin Stejerean
* Maria Martini
* Lukas Krasula
* Mike Colligan
* Margaret Pinson
* Jakub Nawała
* Dawid Juszka
* Mikołaj Leszczuk
* Steve Goering
* Shirin Rafiei

“No Reference Metric Research Paradigm”

By: Margaret Pinson

File: VQEG\_NORM\_2020\_113\_No\_Reference\_Metric\_Research\_Paradigm

The *challenge dataset* idea—a dataset of videos containing just a single experiment.

<https://github.com/NTIA/NRMetricFramework>—a GitHub repo with related implementations.

Pablo suggests adding sigmoidal (or other non-linear) mapping to RCA parameters (when mapping them to MOS). E.g. ITU-T G.107, G.1070, etc:

Q = Qmin + (Qmax - Qmin) \* (R + 2\*R\*(R-0.6)\*(1-R))

Where Qmin=1, Qmax=5, Q=MOS (1-5), and R = 1 - ∑(𝑤\_𝑝 𝑥\_𝑝 )

AGH declared to provide support.

Ioannis

Let’s consider SVR as something that may bring improvement as compared to plain linear mapping.

Steve Goering is working on a larger release of features and models (some previously presented at VQEG) that could be useful for NORM

“Video quality metadata in compressed bitstreams”

By: Ioannis Katsavounidis

File: VQEG\_NORM\_2020\_130

Proposal, standard video quality metadata payload: video quality metric, version, raw score, MOS, confidence interval, scaling method, temporal reference, aggregation method, generation index (see presentation for details).

NR metrics needed:

* camera front end (camera metrics can help)
* Legacy videos
* Video broadcasting applications (transmission over noisy channel)
* Different (non-transcoding) image & video applications

But metadata + FR metric solution would work for most modern transcoding pipelines

* Very little space, need standard format, hardware & service providers have a lot to gain

Ioannis suggests to organise a separate call on the topic. (Maybe in Jan 2021?)

Discussion: Audio call in January

Topic: adding standard video quality metadata to video compressed bitstreams, based on FR metrics. The goal is to enable quality testing at all points in the video delivery network, as an alternative solution to NR metrics.

User Testing during COVID-19

“Precaution for lab experiments during Covid-19”

By: Kjell Brunnström

File: VQEG\_UserTests\_2020\_134

“User test at scale during the pandemic”

By: Patrick Le Callet, Anne flore Perrin, Babak Naderi

File: VQEG\_UserTests\_2020\_138

Break-out Sessions

Questions: <https://docs.google.com/document/d/1pDtPn-esNDjzdzdUGCiC2vEsJq9QSQR-/edit>

AVHD Session

“More Ecologically Valid Subjective Experiments”

By: Lucjan Janowski

File: VQEG\_AVHD\_2020\_115

Open question: what important quality influencing factors can you think of (focusing on the entertainment scenario)?

Alex Raake offered to provide few ideas on that offline.

Likewise if it comes to Niall Murray.

AVHD-AS/P.NATS Phase 2 Project—Standardization and Model Performance Report

By: Shahid

File: (ftp://vqeg.its.bldrdoc.gov/Documents/VQEG\_Stockholm\_Dec20/AVHD-ASP.NATS\_Phase\_2\_Model\_Performance\_Report\_v02.pdf)

Decision: The AVHD/PNATS2 validation report is approved.

A Hitchhiker's guide to SSIM

By: Alan Bovik

File : VQEG\_AVHD\_2020\_149

Al shared the pre-preprint of the paper. Comments and questions are welcomed.

A Comprehensive Analysis of Crowdsourcing for Subjective Evaluation of Tone Mapping Operators

By: Ali Ak, Abhishek Goswami, Wolf Hauser, Patrick Le Callet, Frédéric Dufaux

File: VQEG\_AVHD\_2020\_142

Control group vs “in-the-wild” group (both in the crowdsourcing setup?).

“On a QoE Metric for Live Media Streaming Applications”

By: Rohit Puri (Twitch)

Live streaming (Facebook live, periscope, twitch, youtube live) have tight service availability and latency constraints, large infrastructure investment

Multiple actors including network, media processing, broadcast, playback devices shape experience of a live media streaming audience.

QoE viewer determined by quality of audio, video, sync, latency, …

Are you experts interested in measuring QoE for live streaming application? Yes, this is a natural follow-on to the AVHD-AS / P.NATS Phase II work.

Interest?

* Margaret Pinson (NTIA)
* Rohit Puri
* Shahid
* Alex R. / TUIL
* Q14 work related to this topic: P.BBQCG, P.1204, G.CMVTQ, maybe also P.NATS Phase 3
* Nabajeet Barman (KU, London) [all my previous work including PhD thesis have focussed on live streaming applications with a special focus on Twitch]

ToDo: A dedicated call to be announced for this work item on VQEG email reflector (Shahid/RohIt)

ToDo: AVHD page needs to be updated with the information of the new work item (Shahid/Margaret)

**Need to schedule time for this discussion at the next VQEG meeting**

Decision: Agreement to start this as a sub-project within AVHD

Agreement to hold an audio call

Must coordinate with CGI, due to overlap of interest

Immersive Media Group

“Comparison of Subjective Methods for Quality Assessment of 3D Graphics in Virtual Reality”

By: Yana Nehmé, Florent Dupont, Jean-philippe Farrugia, Patrick Le Callet, Guillaume Lavoué

File: VQEG\_IMG\_2020\_147\_YanaNEHME

Compared ACR-HR, DSIS, and SAMVIQ for graphics in virtual reality

Concluded in favor of DSIS

See publication for more details.

Yana Nehmé, Jean-Philippe Farrugia, Florent Dupont, Patrick LeCallet, and Guillaume Lavoué. 2019. Comparison of subjective methods, with and without explicit reference, for quality assessment of 3D graphics. In ACM Symposium on Applied Perception 2019 (SAP '19).<https://doi.org/10.1145/3343036.3352493>

Yana Nehmé, Jean-Philippe Farrugia, Florent Dupont, Patrick LeCallet, and Guillaume Lavoué. 2021. Comparison of Subjective Methods for Quality Assessment of 3DGraphics in Virtual Reality. In ACM Transactions on Applied Perception (TAP ‘21) - Will be published very soon (in january probably)

Minutes approved.