VQEG Spring 2022 Meeting Minutes

Zoom Meeting link: <https://insa-rennes-fr.zoom.us/j/91993932172?pwd=SnhOYzBld3Z1YnRoSTRPdmJtOEs2QT09>Meeting ID: 91993932172, password: 066770

Day 1: May 9th, 2022

Session 1 — Projects Overview

Projects can be found here: <https://www.its.bldrdoc.gov/vqeg/projects-home.aspx>

AVHD

Not so much activity since the last meeting.

PsyPhyQA

No updates at the moment..

QAH (Quality Assessment for Health Applications)

Slides: [VQEG\_QAH\_2022\_.pptx](https://docs.google.com/presentation/d/1qOfcD-LIVWiPhV2QTtUvloX-Vkl78TqH/edit?usp=sharing&ouid=103256206714529046132&rtpof=true&sd=true)

* Topical review submitted to SPIC in February (objective QA for medical imaging)
* Special session accepted for IEEE ICIP next October in Bordeaux
* Paper submitted to ICIP on QA through detection task of covid-19 pneumonia
* VQEG meeting organization in Rennes :-)
* References:
	+ <https://pubmed.ncbi.nlm.nih.gov/34225264/>
	+ <https://ieeexplore.ieee.org/document/8463297?part=1>

Contact: lu.ge@insa-rennes.fr, meriem.outtas@insa-rennes.fr, lucie.leveque@univ-nantes.fr

SAM

(meetings on Mondays at 5pm CET; once each four weeks): proposal to update ITU-T Rec P.913, generalized score distribution, cross-lab analysis, FOWR, PC vs. ACR in lab and crowd

* The [home page](https://www.its.bldrdoc.gov/vqeg/projects/statistical-analysis-methods-sam.aspx) of the group.
* Send an e-mail to Lucjan <lucjan.janowski@agh.edu.pl> or to Zhi <zli@netflix.com> to join the group.
* Next call on May 30, 2022, 5:00–6:00 pm CEST. Use [this](https://meet.google.com/ebb-qqfq-bkj) link to join our Google Meet meeting.
* FOWR — (?) Few Observers, With Repeated Rating
* GSD-related work (uploaded to arXiv in 2021 and 2022):
	+ arXiv: [2204.10565](https://arxiv.org/abs/2204.10565)
	+ arXiv: [2202.02177](https://arxiv.org/abs/2202.02177)
	+ arXiv: [2204.07131](https://arxiv.org/abs/2204.07131)

QACOVIA

Publication in Electronics 2022, resistance of DL models to compression, update objectives of the group during this meeting?

CGI

Started in the 2018 meeting in Madrid.

The main focus is devoted to analyzing and evaluating of computer-generated content

So far, we have had 27 presentations.

Research topic:

* Gaming quality assessment methodologies
	+ ITU-T Rec. P.809
	+ Development of GIPS questionnaire to measure interactivity in cloud gaming
	+ CROWDG: Method for assessing gaming QoE using a crowdsourcing approach
* Development of video quality dataset:
	+ CGVDS, [KUGVD](https://ieeexplore.ieee.org/abstract/document/8727887), [GamingVideoSET](https://ieeexplore.ieee.org/abstract/document/8463362), [GamingHDRVideoSET](https://ieeexplore.ieee.org/abstract/document/9422736), [LCEVC](https://arxiv.org/abs/2204.05580)
* Gaming quality prediction models:
	+ **Signal based model:** Nofu, NR-GVSQI, NR-GVQM, DEMI, NDNetGaming
	+ **Bitstream payload based model:** P.1204.3 FHD, Deep-BVQM
	+ **Bitstream header based model:** BQGV
	+ **Planning model:** G.1072, GamingPara, G.OMMOG

Standard Activities (Ongoing)

* ITU-T Work Item P.BBQCG: Parametric bitstream-based Quality Assessment of Cloud Gaming Services

Contact: saman.zadtootaghaj@dolby.com , nbarman@brightcove.com

NORM

Slides: [VQEG\_NORM\_2022-124\_SITI-Updates](https://docs.google.com/presentation/d/17Nzj_rPclfHWdNQw9CVhE2OlfxQjn3hlc2jIjJOlkxY/edit#slide=id.g1271a706c69_2_26)

* Three projects
	+ NR metric development
	+ SI and TI clarification, Goals: make SI/TI future-proof, provide a better encoding complexity metric
	+ video quality metadata standard.
* Monthly meetings (link on VQEG website)

Goal for this meeting: update P.910 recommendation draft, see [T-REC-P.910-202111-I!!MSW-E-SITI.docx](https://docs.google.com/document/d/1pGqvifcoYk_nZ33Q-xnbOMTzQl4wpwgS/edit#heading=h.s8a3itid4k6q)

IGVQM

No updates at the moment.

JEG-Hybrid

New website: <https://vqegjeg.github.io/jeg-hybrid/>

Activities (from website)

* Modeling single observers: to mimic the behavior of single observers in subjective experiments, to investigate/understand their characteristics, and potentially to run “virtual” subjective experiments.
* Modeling VQM disagreement: to achieve better design of subjective experiments, to give hints about which video sequences are most likely to make VQM predictions fail.
* Training machine learning models with a low amount of subjectively annotated data, and/or with (less reliably) automatically annotated data, and using them to improve existing VQMs.
* Improving subjective testing and experiments by finding the most interesting sources (SRCs) and/or processed video sequences (PVSs) to work with.
* Overcoming the MOS limitations by considering subjective score ranges and distributions.
* Developing methods, approaches, frameworks, and tools to promote reproducible research in the context of Video Quality Assessment (VQA).
* Using large scale (not subjectively annotated) datasets to identify shortcomings of common Video Quality Metrics (VQMs) in various conditions, e.g., different coding parameters, lossy channels.

Note: need for Margaret to update the link of the website.

IMG

* Goals: create datasets of immersive media content, provide QoE guidelines, subjective test methods, objective metrics, etc.
* Activities: further data analyses from cross-lab tests with 360-deg videos, ACM SIGMM records (quarterly newsletter): <https://records.sigmm.org/2022/04/11/vqeg-column-vqeg-meeting-dec-2021-virtual-online/>
* Contact: jesus.gutierrez@upm.es, pablo.perez@nokia.com, zzchen@whu.edu.cn

5GKPI

* Goals: define relevant use cases, study QoE aspects for video in mobility.
* Monthly meetings (2nd Thursday each month, 3.30pm CET).
* Main activity: contribution to ITU-T G.QoE-5G.
* Ideas and proposals: write a magazine article, subjective assessment of ToD videos, send video KPIs to the network, move to / rename as 6G?
* Contact: pablo.perez@nokia.com
* Possible liaison with 3GPP SA4. Activity similar to TS 23.501: System architecture for the 5G System (5GS)

Support groups

* eLetter: a little bit dormant for a while… Rebuilt it?

Maybe something with the QAH ICIP special session next October (introduction, abstracts, links to the papers)?

* Qualinet
* Tools and subjective labs setup: <https://vqeg.github.io/software-tools/>
* IRG-AVQA: intersector rapporteur group

HFVE

IEEE standards association working groups on human factors for visual experience.

* IEEE P3333.1.3: DL-based assessment of VE based on HF: approved and published.
* IEEE P3333.1.4: light field imaging, draft standard submitted, not approved yet.
* 3D, UHD, HDR.
* VR and MR.

Session 3 — NORM

Presentation #114, Yiannis Andreopoulos, Domain-Specific Fusion Of Multiple Objective Quality Metrics

* iSIZE
* Solutions: BITSAVE, BITCLEAR, BITGEN
* 3 challenges: objective metrics are myopic (humans as well!), exploration space, metric overfitting?

Questions: - Did you make aware users can see strange things?

- Did you analyze results based on subjective experiments only? Can you spot how much you gain bit rate with using VMAF at all?

Session 3 — CGI

Presentation #113, Jerry (Xiangxu) Yu, Subjective and Objective Quality Assessment of UGC Gaming Videos

* LIVE-YT-Gaming and online study: database
* GAME-VQG specifically designed for gaming videos
	+ Use two modules els to predict the quality:
		- extracted NSS features
		- extracted features pre-trained CNN (DenseNet-201)

→ Youtube project completed, Facebook project ongoing

Q1: did you use crowdsourcing for test?

* it was a remote test because of Corona situation, and still people are recruited from university, so there was control over environment.

Q2: Any comments of the computation complexity?

* we did not analyze it, but we expect to be more complex due to two modules it has.

Q3: Did you analyze the performance across resolutions?

Presentation #118, Nasim Jamshidi, Deep-BVQM: A Deep-learning Bitstream-based Video Quality Model

* Future perspectives: extend codec type, extend resolution ranges, predict non-gaming content.
* Question: Proposed model performs frame level quality prediction. What is the utility of such a model? People probably cannot perceive quality changes within frames.

Session 4 — NORM

Presentation #124, Werner Robitza, Updates on SI/TI

Slides: [VQEG\_NORM\_2022-124\_SITI-Updates](https://docs.google.com/presentation/d/17Nzj_rPclfHWdNQw9CVhE2OlfxQjn3hlc2jIjJOlkxY/edit#slide=id.g1271a706c69_2_26)

* New vs. old SI/TI (HDR)
* Resolution dependency
* Compression efficiency
* P.910 recommendation updates (feel free to add comments!): <https://docs.google.com/document/d/1pGqvifcoYk_nZ33Q-xnbOMTzQl4wpwgS/edit?usp=sharing&ouid=103256206714529046132&rtpof=true&sd=true>

Presentation #120, Lukas Krasula, Banding annoyance vs. overall quality

Banding: CAMBI 8b, compression artifacts: VMAK 4K.

VMAFBA can improve VMAF. Check out: <https://arxiv.org/abs/2202.11038>

Day 2: May 10th, 2022

Session 5 – AVHD

Presentation #111, Mikołaj Leszczuk and Tomasz Konaszyński, Analysis of the Influence of the Experiment Conditions on the Subjective Assessment of the Quality of Video Transmission and Its Stability

Research questions: relationship between subjective assessment and…

* Video sequence variation/repeatability
* “Learning” with views
* Movie order layout

Experiment: 180 video sequences, ACR, 32 participants

Q: Question by Shirin Rafiei about RQ2, the first result plot, what is the scale on the Y axis? Answer: This is the number of ratings (R1>R2, R1=R2, R1<R2).

Presentation #121, Lucjan Janowski, More Ecologically Valid Experiment Design - Two Proposals

Human visual system (HVS), quality of pixel (QoP), quality of experience (QoE).

When asking about quality, experience disappears?

2 proposals:

* ACR without scale: “thinking aloud” psychology method (see the evolution of scoring system, see all quality dimensions) - but data analysis is more challenging, and some testers do not speak.
* Your Youtube, our lab: rare popup with a quality question, user selects the content, real experience (full screen, comments, watch it faster, jump, etc), standard quality scale - but people forget about quality, some people always jump through the video.

Question by Shirin Rafiei: Is the method strongly biased by intro/extrovertism? The moment we ask someone about quality the person responds about quality? Answer by Lucjan Janowski Yes, this is the main threat of the experiment. It is not a QoE (Quality of Experience) experiment, this is a QoP (Quality of Pixels) experiment. Recommendation to run 2-3 people in the experiment to understand better how people distinguish between.

Question by Jakub Nawała: Is more like a comment. Thinking aloud protocol is the way to check if people understood the instructions. The second thing is that Your YouTube Our Lab influences the quality of video. The third thing is that perhaps we should measure behavioral measures. Answer by Lucjan Janowski Yes, these are possible solutions.

Question: What is exactly measured by ACR without the scale. Answer by Lucjan Janowski Yes, this is just an audio wave.

Question: About the different measurement points. Answer by Lucjan Janowski This is not a coincidence; we mainly care about QoP (Pixels). We should just state it clearly in papers. QoP is much more useful and common than QoE.

Presentation #127, Kjell Brunnström, Video quality testing of Video Assistant Refereeing (VAR) Systems

Field tests in Germany and the Netherlands, certification events in Sweden.

Video quality measurements: 25 video experts, ACR-HR, 1080p, 1080i, 540i.

VQM VFD better (PSNR worst). Yet VMAF was used.

Questions: Meaning of yellow in objective models evaluation table? VMAF compared to RMSE values, statistically significantly different based on RMSE. Codec fixed (H.264), maybe consider using bitstream-based metrics (e.g., P.1203 or P.1204)? Hopefully later ;-) good suggestion.

Questions: Differences with video experts compared to “normal” subjective tests? Can we compare with naive viewers? Good suggestion, long range of bitrates they couldn’t notice either, slope then it dropped close to 20Mbps.

Question: Why did you decide to use an interlaced format? Football: broadcasted!

Session 6 – IMG

Presentation #129: Gunilla Berndtsson, On the advances of the ITU rec. on QoE Assessment of eXtended Reality Meetings (P.QXM)

Quality of Experience in Telemeetings and Videoconferencing: A Comprehensive Survey. IEEE Access. Janto Skowronek et al. 2022. To be published soon.

Presentation #117: Zhengyu Zhang, Deep blind light field image quality assessment by extracting angular and spatial information

Not enough LFI data to train a CNN model, and no CNN model designed for LF-IQA.

Proposed metric: DeeBLIF (no-reference metric): angular-spatial patch generation, 2-stream CNN model (angular stream, spatial stream).

Experiment: Win5-LID dataset, PLCC, SROCC, RMSE.

Question: Are subjective ratings included in the dataset? Ground truth for each image.

Comment: Would be interesting to validate the metric on other datasets.

Question: Why 5x5? To reduce computational complexity.

Presentation #125: Marouane Tliba, Representation learning optimization for 3D point cloud quality assessment without reference

Session 7 – JEG-Hybrid

Presentation #103: Enrico Masala, Updates on JEG-Hybrid (new website)

Brief update about the activities going in the group in the last years, which gradually evolved over time to include several areas of Video Quality Assessment (VQA) in addition to the original aim to develop a no reference hybrid/bitstream model. Currently the group is not directly seeking the development of new metrics or tools readily available for VQA.

To reflect the evolution that happened in the last few years, the JEG-Hybrid website has been redesigned and put on GitHub pages. Such a setup should also facilitate maintenance and timely updates. The link to the new JEG-Hybrid website is:

<https://vqegjeg.github.io/jeg-hybrid/>
The new website will include a static mirror of the old one, not to lose the history of the group activities.

The link should be updated on the VQEG website, in the page describing the JEG-Hybrid project.

Presentation #104: Lohic Fotio Tiotsop, On the Sensitivity of Artificial Intelligence-based Observers to Input Signal Modification

Lohic Fotio Tiotsop presented recent activities going on in the JEG-Hybrid group, in particular modeling the behavior of single observers in subjective experiments through DNNs, currently investigating the stability of the performance of such DNNs.

List of related publications is available on the new JEG-Hybrid website here: <https://vqegjeg.github.io/jeg-hybrid/publications> where download links for all publications are available (either the final version submitted by the authors or the editorial version whenever possible).

Session 8 – QACOVIA

Presentation #115: Alban Marie, Correlation of quality metrics with AI accuracy

Question by Lucjan Janowski if there is a chance to compare to the original image. Answer comparing distorted images (Full Reference).Maybe I am missing the question, but we can compare prediction between on undistorted and distorted image by using a distance function between both model outputs.

This is what is actually done on "Image level" experiment where KL-divergence is used as the AI performance metric (y-axis on plots)

Question: What is the task? Answer data classification and segmentation.

Question: are any other metrics used? Answer: No Reference metrics will be used.

Question: Is the noisy graph with correlations coming from the same content? Answer: yes. Comment: maybe some content is not affected by compression.

Question by Mikołaj Leszczuk: Or rather comment, this presentation is a nice match to the presentation of Mikołaj Leszczuk to be held on this Thursday.

Presentation #134: Lucie Lévêque, Are facial expression recognition algorithms reliable?

Comment from Lucjan Janowski: looking forward to seeing results from humans, without distortions.

Question: Is there an intensity level of emotions? Answer: in most datasets there is only one emotion as ground truth.

Question: Maybe it is better to detect some key points on the face? Answer: yes, there are systems like that (FACS = facial action units systems).

Question by Irene: How can confidence be used? Answer: Good question, maybe as a weight?

Question by Irene: Maybe it would be beneficial to add more than one emotion? Answer: yes, but matter of time…? And ground truth / datasets only contain one emotion.

Question by Lucjan Janowski: The database contains only faces or other things as well? Answer: Only faces but there were some other pictures removed.

Session 9 – SAM

Presentation #133: Andreas Pastor, Improving maximum likelihood difference scaling method to measure inter content scale

Question/Recommendation by Dietmar Saupe: You are relying on a normal distribution for the simulations and use MLDS as the solver but MLDS has its own assumption. I recommend you to try this out with the probabilistic model from MLDS and to check if your convergence goes to zero or not.

Presentation #107: Jakub Nawała, Generalised Score Distribution: A Two-Parameter Discrete Distribution Accurately Describing Responses from Quality of Experience Subjective Experiments

Here is the arXiv handle to the preprint of the paper: <https://arxiv.org/abs/2202.02177>

Question from Irene: is the code available? Yes, GitHub repository. <https://github.com/Qub3k/subjective-exp-consistency-check>

Question from Irene: What kind of statistical tests can be used after the model is applied to data? Could be used as a base to create a test.

* Since the GSD describes the subjective data well, it can be used as a basis of parametric statistical tests. Those would be more powerful than nonparametric tests used currently. However, we need to first create such parametric tests. Stay tuned! :)

Question from Dietmar: Why is the GSD so special? There are plenty of two-parameter continuous probability distributions that could be mapped to the 5-level ACR scale. Why not try to use these?

* We wanted to use a completely discrete model. That is, a model that is *not* mapping from a continuous domain to a discrete domain. We wanted to do so, because we believed that this would provide a benefit over models that map from a continuous to a discrete domain.

Question from Dietmar: How about reparameterising the Ordered Probit distribution to fit GSD’s characteristics. Would that work?

* Yes, I guess this would work. However, I am not sure if parameter interpretability (of the reparameterised Ordered Probit model) would be retained though.

Suggestion from Irene: You should consider creating handy packages implementing the GSD. For example, an R package would be nice.

Presentation #132: Ali Ak, Spammer Detection on Pairwise Comparison Experiments: The Impact of PoE Scenario

PoE: preference of experience.

Be careful with Amazon Mechanical Truck crowdsourcing platform!

Comment from Lucjan: very important to do outlier detection.

Presentation #140: Dietmar Saupe, Subjective image quality assessment with boosted triplet comparisons

<https://ieeexplore.ieee.org/document/9559922>

Presentation #136: ​​Chama El Majeny, Yvann Gouraud, and Liu Jiawen, Description & Function of Subjective test analysis tool

The code will be available in roughly two weeks from now. This means the code would be available around May 24, 2022.

Day 3: May 11th, 2022

Session 10 – ITU Activities

Presentation #142: Chulhee Lee, Alexander Raake, Overview of ITU activities

Q10/12: Conferencing and telemeeting assessment

Performance of P.1204.4 on AV1 Encoded Video: <https://www.itu.int/md/T22-SG12-C-0008/en>

In case of any questions regarding ITU-T SG12: martin.adolph@itu.int

Session 11 – IMG

Presentation #116: Shirin Rafiei, Augmented Remote Operating System for Scaling in smart mining applications: Quality of Experience and User Experience aspects

Ongoing investigation. Mine tools occlude views, operators need perspective views.

AROSS method. Experiment 1: mixed interface. Experiment 2: single interface.

No sig difference between disocclusion and novel perspective, but higher MOS for disocclusion augmentation (disocclusion beneficial for user experience).

Question from Lucjan: If I detected using left original, then right original, then I detected the same? Answer: mine wall divided into 3 parts, thus question asked regarding the right part. Lidar view: no color, thus need for validation.

Question from Jakub: About the MOS, based on which question? Answer: one question was related to “how could you control the robot arm when you have this active interface?”. Original view okay, but disocclusion: 100% transparent arm. Likert scale.

Other questions asked to participants were about how you can control the arm, and about depth estimation.

Presentation #130: Pablo Pérez & Jesús Gutiérrez, Follow up on the new test plan

Declaration + question from Lucjan: very interested in being involved! About optimization… Go to the UX community? Necessary to have some people who really work on it?

Question from Kjell to Lucjan: what do you mean about optimization? → To be able to control the system in the future (Lucjan). Jesús: target continuous bodies.

Question from Irene: very different systems, very different contents; difficult to define requirements / to compare the systems. How do we evaluate this? What task to use?

Question from Jakub: You talk about communication tasks. What about tasks that require only one person (e.g., exploring some VR space or using AR glasses to aid repair or service tasks)? Is this also of interest?

* Jesus, would you be so kind and address this question?

Presentation #131: Ali Ak, The Effect of Temporal Subsampling on the Accuracy of the Volumetric Video Quality Assessment

Question from Irene: how big was the group of pictures in V-PCC random-access? Answer: Hard to tell, Ali will answer later.

Question from Shirin: Really close numbers, how do you make a difference? Answer: normally statistical analysis, here we didn’t.

Question from Irene: Did you try other metrics like PCQM or point SSIM? A lot of calculations.

Question from Aladine: More complex? We are comparing PCA for selecting a view around the point plot to get the best representation.

Presentation #135: Waqas Ellahi, A machine-learning framework to predict TMO preference based on image and visual attention features

Question from Aladine: How can we compare more than 2 images? How to extend your work? Answer: For ML, first we need some datasets. We have some techniques to rank more than 2. Not possible with SVM.

Advertisement: Irene Viola

1st international workshop on interactive extended reality (IXR 2022)

Deadline: 20th June

<https://ixr2022.itec.aau.at/>

Session 12 – QAH

Presentation #122: Meriem Outtas, Image Quality Assessment Through the Detection Task of COVID-19 Pneumonia

Question/comment from Lucjan about denoising… Maybe try with students instead of AI…

Presentation #125: Marouane Tliba, Representation Learning Optimization for 3D Point Cloud Quality Assessment without Reference

Question from Andreas on the explanation of the training procedure: you have 2 tasks? Answer: one task. We also want to train our encoder.

Comment from Lucjan: To understand an example, how it was invented = for crime. Not biased by other elements.

Presentation #106: Yuhao Sun, Visual Distortion on Medical Images: NR-IQA Evaluator in CT Scans

Question from Ali: How do you distinguish the contrast used for enhancement and contrast considered as distortion? (the answer is given throughout the presentation)

Question from Lucjan: In the pre-test questionnaires, it is mentioned that the compression is one of the 4 types of distortions encountered in medical imaging. However, it is a common practice to not to compress medical images, how come this can be encountered?

Question from Lucie: Do clinicians and radiologists know what is blur, contrast, etc distortions? Answer: Examples were shown to them.

Session 13 – NORM

Presentation #102: Hadi Amirpour & Vignesh V Menon, VCA: Video Complexity Analyzer

VCA is open-source and available on the link below, we are open to contributions.

<https://github.com/cd-athena/VCA>

Question from Ali: What about HDR videos? Just like SI-TI, can we also extend VCA similarly? Answer: adding new features for instance, like brightness. So it can be extended in the next versions.

Question from Ali: How does the SI-TI and VCA correlate? Answer: probably don’t correlate that much because VCA correlates well with encoding bitrate whereas SI/TI does not.

Question from Shirin: Real-time streaming, overlay. Detect the complexity independently…..

Question from Jingwen: How do we calculate temporal complexity? + add motion

Question from Jingwen: at slide 18,  why do we have a high complexity peak at the left side of the plot? It is due to a scene cut in the video. The example of the scene cut above and the plot below are not related, sorry for that confusion. They are two separate examples.

Presentation #105: Hadi Amirpour & Vignesh V Menon, VCD: Video Complexity Dataset

Link to dataset: <https://ftp.itec.aau.at/datasets/video-complexity/>

Question from Andreas: Did you use more modern codecs? answer: we are working on it currently, stay tuned

Question from Jingwen: mentions a specific case of scene cut and considering marking/detecting such scene cuts. Answer: it is complex and detecting those cases would require manual intervention probably

Comment from Enrico: additional sequences could be taken from the SRC of the JEG-Hybrid large scale dataset: ftp://[ftp.ivc.polytech.univ-nantes.fr/VQEG/JEG/HYBRID/hevc\_database/](http://ftp.ivc.polytech.univ-nantes.fr/VQEG/JEG/HYBRID/hevc_database/), (no password necessary to access the ftp) described here: <https://vqegjeg.github.io/jeg-hybrid/resources/large_scale_dataset_HEVC>

Presentation #109: Mikołaj Leszczuk, User-Generated Content (UGC)/In-The-Wild Video Content Recognition

Suggestion from Pablo: Take a look at the paper from last week’s (2–6 May, 2022) ACM CHI conference. Something about taking into account dissenting voices in data labeling.

[Mitchell2022] Mitchell L. Gordon, Michelle S. Lam, Joon Sung Park, Kayur Patel, Jeff Hancock, Tatsunori Hashimoto, and Michael S. Bernstein. 2022. Jury Learning: Integrating Dissenting Voices into Machine Learning Models. In *CHI Conference on Human Factors in Computing Systems* (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 115, 1–19. <https://doi.org/10.1145/3491102.3502004>

Suggestion from Nabajeet?: If the dataset can be shared per request basis, the copyright issue can be avoided. answer: yes we can do that, please let me know if you want to have access to the dataset.

Question from Ali: You said you worked per shot basis but in some of the in-the-wild / professional mixed contents, it can be difficult to detect changes in shot, especially for cases where the news studio has a large overlay and the large screen in the studio displays a video with scene cuts.  Are they common enough to care about? If so, how do you treat such cases? Answer: Such cases are not that common in the dataset we used. The videos used in this work have clear classifications as either professional content or in the wild content.

Question from Andreas: Did you use any other indicator? We didn’t use all the indicators. Interlace for example was a highly correlated indicator to identify professional content but we excluded it.

Presentation #110: Mikołaj Leszczuk, Advanced Visual Quality Indicators

We are working on a python package. We also work on a web service that can be used in a drag and drop fashion for various video quality indicators such as SI/TI (and their potentially better counterparts), etc.

Maria Martini said that her team has implemented video indicators that can help characterize video content.

Presentation #138: Margaret Pinson, Proposed revisions to ITU-T Rec. P.913

<https://docs.google.com/document/d/1of8ZsB2nZpi8fn5AOD5r9BfgjuGI4X7L/edit>

Comment from Nabajeet: in passive/spectator gaming, we use 30 seconds videos in general....we can probably add that too.

Comment from Saman: You can add P.809 recommendation as reference.

Day 4: May 12th, 2022

Session 14 – AVHD

Presentation #139: Olof Lindman, Fredrik Lundkvist, Swedish Television (SVT) Open Content

<https://www.svt.se/open/content>

Ok to share them on cdvl.org

Question from Anne-Flore: are you planning to extend the number of content? Short answer = yes. Time frame depends on a lot of things. Being discussed. Much easier to motivate internally if we show it is so, e.g., please send an email to them if you use them for something, so it is easier to make a point for that mething people will use.

Presentation #128: Babak Naderi, On subjective evaluation of video quality with the crowdsourcing approach

<https://github.com/microsoft/P.910>

<https://arxiv.org/pdf/2204.06784.pdf>

arXiv ID: [2204.06784](https://arxiv.org/abs/2204.06784)

There is a chance that Babak and Ross will make the related subjective responses available. This may be interesting to Jakub and Lucjan (in the context of their work on the GSD—see presentation no. 107 above)

Babak asked for proposals of data sets that he could use to compare crowdsourcing and lab experiments. Lucjan’s proposal is to use the data set generated by Lukas Krasula (see [presentation no. 120](https://docs.google.com/document/d/1anUTmFPQGmamA77_IHq66gwnWa76dYoYCBIgRvoMIrM/edit#heading=h.qr0xlvnclrpv) above).

Presentation #137: Pierre David, Lucjan Janowski, and Kjell Brunnström, Social Media Video Quality: A Three-lab experiment

All videos were upscaled to fill up the entire screen (mobile and PC).

Margaret asked about the related paper. There is no one yet. However, there is a plan to write one.

Session 15 – QACOVIA

Presentation #108: Mikołaj Leszczuk, Objective Video Quality Assessment Method for Face Recognition Tasks

Use PIQA (?) Python PyTorch implementation of NR metrics, since it’s much faster than their Matlab implementation.

Session 16 — Future On-line (?) Meetings

Let’s schedule a follow-up on-line call to conclude this discussion.