VQEG Meeting, San Diego

February 29 to March 3, 2016

## Monday

ICDM and VQEG have potential overlap of interests, to standardize on HDR viewing equipment.

**Action Item:** Rename HDR to HDR/WCG (High Dynamic Range and Wide Color Gamut ), agreed to at the last meeting

**Action Item:** change “Active Projects” to “Working Groups” on VQEG website.

**VQEG Progress Report:** articles are **due March 18**. Kjell is organizing the progress report in a google document.

**Proposal** to use Google docs for meeting minutes, presentations, and similar documents. Then download the final version to VQEG ftp server.

Monday morning schedule will be left as-is. Presenters are encouraged to give a “teaser” summarizing their future session discussion.

# 3DTV

Presentation:

Naeem Ramzan, “Performance evaluation of Multiview HEVC”

File: VQEG\_3DTV\_2016\_129\_performance\_evaluation\_of\_multiview\_HEVC

# 3DTV Project

**Note:** All of the co-chairs agreed to close this project.

**Proposal:** shut down existing 3DTV; move remaining work items within new immersive media group; ask for volunteers to lead the effort. This will be discussed during administrative time later this week, after talking by email among co-chairs not present.

**Agreement was reached** to accept this proposal. 3DTV will be closed, and “find subjective methods” task (1st bullet point) of 3DTV will be moved under immersive media group as a project.

# Immersive Media Group

**The interested parties will take the following text and write a revised group mission this week.**

Revise subjective assessment methodologies to evaluate Multiview technologies, including full parallax.

How to measure the results of light field processing also called plenoptic, such as interactive refocusing or changing point-of-view.

Baseline quality assessment of today’s systems

* Using repurposed traditional content for virtual reality
* New content captured specifically for virtual reality
	+ 360 cameras
	+ Light field cameras
* Virtual reality gaming

Display requirements (as per ITU-T Rec. P.914)

How to assess tradeoffs, particularly when comparing very different devices (e.g., a low resolution virtual reality system versus a high quality 4K monitor).

Investigations of both passive and active systems

Systems with and without feedback (in response to the viewer’s actions)

Augmented reality

Viewpoint responsive (e.g., can look around a tree, virtual reality gaming)

View direction responsive (e.g., can move your head)

Non viewpoint responsive (e.g., traditional stereoscopic 3DTV)

Displays ranging from small devices to theater screens

People interested in co-chairing new group:

* Glenn
* Kjell
* Phil
* James
* Patrick
* Zhenzhong Chen very interested in co-chair; or leading a project on quality assessment for virtual reality

**Action:**  send liaison to ICDM on the formation of this new group; ask for involvement and consultation from display technology manufacturers. Kjell and Jorge will take the lead on this liaison.

**Assessment of 3DTV group:**

1. Define suitable methodologies for subjective quality assessment of stereoscopic 3D video: **(Incomplete)**
2. Investigate the influence of viewing environment, test set-up and display equipment on subjective quality **(Done)**
3. Objective video quality metrics for stereoscopic 3D **(wait for proponent to propose)**
4. Analysis of frame compatible 3D video format representations **(Withdrawn)**

Issue to be discussed later this week, during the admin time (8:30am to 9:00am)

**Presentation:**

Jorge Caviedes (Arizona State University), “Meeting operational objectives”

File: VQEG\_VIME\_2016\_114\_Operational\_Objectives

**Presentation:**

James Goel (Qualcomm), “Proposal for low impairment visual quality assessment group”

File: VQEG\_ADMIN\_2016\_127\_proposal\_for\_new\_VQEG\_group

**Chair: Phil & Naeem & Patrick**

**VESA Liaison: James**

Qualcomm moves to create new group.

**Approval for new group “Visual Lossless Quality Assessment”** This group will use the JEG reflector

People interested in working on this subject:

* James, Chulhee, Phil, Patrick, Naeem, Alexis, Jorge, Ioannis, Glenn, Kjell, Ludo, Zhenzhong, David Hoffman (Samsung)

VQEG will try to develop a charter and group name and kick-off meeting date this week. This will be discussed Thursday.

## Tuesday

# Minutes Tuesday March 1st 2016

New Arrivals Today

Louis Lee from Qualcomm responsible for video processing

David Hoffman from Samsung display

Anil Kokaram and Chao Chen YouTube/Google video infrastructure

Keith Lee AMD Canada

Jim Hunkins AMD Canada

Minutes from yesterday finessed

HDR-WCG is new name for the HDR group

3DTV group closed and remaining work items incorporated into Immersive media group.

Immersive Media Group agree to revise the charter for this group later this week

Minutes from yesterday seconded.

9:06 am

Presentation from Michele Saad @ Intel  [Ramesh Jaladi, Philip Corriveau]

“Viqet and Subjective Response to PostProcessing Effects”

From VIME group (no ref quality assessment for photos)

Used 16 scenes and enhanced sharpness even beyond the point its over sharpened

Measured subjective response SS-ACR study 300 participants

Sharpness measure vs MOS .. showed MOS versus log sharpness. But depends on the scene.

INSLA Adjusted Subjective Data vs VIQET score gives 0.83 rank order correlation (SROCC) against MOS. No overlap between training and testing sets.

VIQET Android App available on github under vqeg project

Next meeting of Vime TBD

Question from Yiannis about whether this study incorporated HDR or not .. no is the answer

Question on resolution etc of the study: was crowd sourced online with varying resolutions of pictures, but with some minimum picture size.

Anil Kokaram @ YouTube on Audio Video Quality at YouTube

YouTube provide 2k, 4k, 8k, HFR video, VR 360 stereo-3D up to 2015

YouTube care about new formats such as HDR, VR, 8K

Scale is important: latency, reliability, bandwidth, adaptive bitrate

YouTube also care about quality measurement: QoE and video-audio quality

YouTube video pipeline is responsible for processing all the uploaded videos

Measurement of video quality is used to decide if a video needs to go through pre-processing before transcoding

Ingest trend to YouTube:

* 1080 and 720 dominate the ingest.
* 4K and 8K is growing rapidly.
* As bandwidth increase, more and more high bitrate videos are ingested to YouTube
* Codec-wise, x264 dominate the ingest. Flash is disappearing.
* 90% of content is progressive.

YouTube detects interlaced videos

high resolution videos have more progressive versions than low resolution videos.

The distribution of ingest in the figure is in the unit of video length

An example of subjective study, we use subjective test to decide quality saturation bitrate of 4K, choosing dashs band for 4K, calibrating metrics for banding artifacts

YouTube employ machine learning methods for predict optimal preset/encoding parameters for video clips at low cost.

HDR in YouTube:

* youtube = VP9 profile 2 10bits
* Advocating HLG and PQ for all compliant TVs
* Metadata signaling for HDR delivery
* Check the reliability of TV display for HDR

QA:

* YouTube expect to use quality metrics to optimize watching time.
* ICTCP color space is not yet supported by YouTube
* YouTube has not yet started VR/AR quality assessment.
* YouTube plan to release video quality assessment database upon the agreement of creators

Comment: measurement of sharpness should be applicable for all media, SDR, HDR, VR …

**HDR Kicked off 10:35 am**

Presentation from Patrick Le Callet co-chair of HDR @ VQEG

New metrics for HDR-Video being worked on in MPEG; PSNR and others even with simple PQ style mappings don’t work apparently. These new metrics are only available for MPEG and ITU members.

Question about how to make MPEG HDRTools available to VQEG? Apple may not wish these to be available outside its original scope.

**Question:** How to push HDR-VDP2.2 into a standard for HDR objective image quality assessment?

**Proposal:** VQEG writes a technical report and send it by liaison to JPEG, MPEG, ITU and VESA.

**Patrick Le Callet HDR-WCG Project**

Update: working on experimental method, considering various scenarios, focus on artistic intention, include both MOS and visual intention

Tone mapping progress:

* studied impact on tonemapping on HDR compression
* Comparison several tonemapping operator with single-exposure images in presenting HDR scene
* Method for HDR STUDY content selection

Studies and databases:

* Eyetracker database for HDR
* TMO comparison data base

Objective measures for HDR content:

* HDR-VDP2.2, available to VQEG
* contributing to MPEG through Samsung

Next step: extending objective measure profile

**LUNCH**

Prof. Patrick Le Callet

Video Quality assessment of HDR content

Many dimension for select a video to test, different content, different distortion, color information, dynamic range….

QoE as video quality: definition: degree of delight or annoyance of the user of an applicaiton or service

QoE results from the fulfillment of user’s expectations

Image/Video quality is not QoE, but one component of QoE.

a model of 3D QoE, based on 2D Image Quality, depth quality, visual comfort,

But QoE should also include expectation, cost, context

* viewing condition: If following BT500 recommendation, for peak luminance and environmental luminance, we cannot show HDR
* display: there is no true HDR, display always use TMO
* content selection: many distortion dimension to consider
* Methodology: using or not using a HDR as reference in evaluating TMOs may affect MOS.
* New QOE model: consider artistic intention, which may change according to viewing environment. TMO may also affect attentions (measured by saliency map)

Image quality metrics:

a perceptual based approach (try to model HVS) for VQA.

1. map reference image and distorted image to visual units according to distance and model of display
2. Take difference and normalize the difference by visibility threshold for each frequency subbands
3. Pooling the normalized errors to single score.

Extension to HDR, use emitting luminance as display model and then transformed into perceived luminance

HDR VDP 2.X: good at predicting JND, not good for video, complexity is high

HDR-VQM, first use emitting luminance as display model and then transformed into perceived luminance and the apply gabor filtering to extract spatial-temporal features and perceptual errors. Then apply spatial-temporal pooling.

Current Progress in HDR Subjective Studies

VQEG groups need a consumer grade display

Schedule: Mid April, first round of subjective evaluation completed (pending display availability)

Request for colour volume conversion algorithms for testing HDR->LDR and vice versa.

Rec 709

Using “SAMVIQ”

Target end of April

Tuesday afternoon

The ULTRA UHD session was moved to Thursday afternoon to accommodate discussion on HDR-WCG.

HDR-WCG session contd…

Patrick continued his presentation on HDR-WCG

The participants that are interested in contributing to scenario 2, which is 4k HDR (see slides for scenario details):

* Acreo, Qualcomm, Yonsei, Youtube, University of Nantes, Opticom, Samsung, Netflix, Ghent University, Sky, and UWS.

The HDR-WCG session finished at 15:00pm

Coffee break at 15:00 -15:30 pm

**AVHD session**

Chris chaired the AVHD-AS project synopsis.

The deadline to indicate preliminary interests in participation in this AVHD-AS project is at this VQEG meeting. Three interests received so far. There is a second deadline to confirm interests to be a proponent after finalization of all documents.

Changes made to the AVHD-AS project synopsis

\*Removal of the hybrid models from the AVHD-AS project to avoid duplication of work. Discussion to continue offline.

\*The duration of the sequences will be between 1 minute and 5 minutes

\*Inclusion of 720p@60fps

\*Removal of yuv 422 and yuv 444.

\*Acceptable viewing devices - monitor and TV are equivalent. Smart phone viewing devices are included in phase 1. Notebook viewing devices is under discussion for phase 2.

\*to mention "for the mobile resolution, the display resolution may not match 1080p" in subjective experiments. (this text will be added to the subjective experiment document later this week)

\*The monitor is not expected to be a variable.

Shahid presented the subjective test conditions document.

The low impairment compression charter discussion was moved to Wednesday.

## Wednesday

**IRG AVQA Session:**

SG9 report by Chulee:

SG9 finished the 3D display requirement

Fatigue related aspects

There will a recommendation in approximately two month s

P.913 is revised and revision was approved

revised P.912, consented at last January

G.QUHD for UHD TV signals, delayed until next year.

J.OptimalTransmission, could be delayed

VQM\_HEVC: applicability of existing recommendation, if so we can extend, otherwise we can ask for a new recommendation

J.343 revised (speed up)

see details in the SG9 Q2/Q12 session below

SG9 last meeting of this study period will be this August. Geneva

Christian SG12:

Mainly Q14 work, good progress for validation procedure

Good progress for specifying phase 2 of PNATS, end of April a detail test plan should be available

Q13 started a call for participation of G.OM\_HEVC (an opinion or planning model for HEVC)

ITU-R WP6C: Chulhee

The dominant issue is HDR, ITU-R produced 5 draft new Recommendations and reports, and also they opened new questions. There was proposal for subjective testing of UHD TV.

IRG AVQA session closed ---------------------------------------------------------------------------

**Rapporteur Meeting of SG9 Q2 and Q12 --------------------------------------------------------------------**

Chulhee:

P.3D already finished (display requirements for 3D video quality assessment) TD??

P.av-ims Immersive subjective testing method for audio video and av stimuli

M: VQEG has something to contribute something there but progress is slow

J.vqm-hevc objective perceptual video quality measurement methods for h.265

Hybrid model extension for H.265 is discussed. Chulhee proposed to validate the existing models for hevc, and if models do not work well organize a new work item

to submit new work item. Producing databases (realistic transmission errors) is a bit tricky in this case.

Ioannis: doesn’t see the practical application, its all HTTP now.

J.q-uhd: Quality measurement for UHD services. need a lot of preprocessing, as quality of HD or 720p can be higher than UHD. So mainly we need some tool to find out if a certain signal justifies to be used in UHD

J.343 revision is necessary to increase the speed.

Rapporteur meeting formally closed ----------------------------------------------------------------------

**Decision of the next meeting ---------------------------------------------------------------------------**

decided venue: Sky (London)

decided date: 24-28 October 2016

options for next meeting: Huawei (China), Krakow (AGH), California (Netflix)

---------------------------------------------------------------------------------------------------------

**VQEG JEG Hybrid -----------------------------------------------------------------------------------------**

Glenn´s presentation on how JEG works

Enrico's presentation. on agreements/disagreements of video quality metrics

Ahmed's presentation. on content features

---------------------------------------------------------------------------------------------------------

**MOAVI**March 2nd 1:30PM

Mikołaj Leszczuk (AGH) & Lucjan Janowski (AGH): MOAVI: Findings on NR quality indicators [file=VQEG\_MOAVI\_2016\_108\_NR\_Quality\_Indicators]

Signal-Based, NR Indicators for Artefacts of Various Origin:

1) H.264 quality indicators
Merging Exposure and Brightness
Frame Drop Detection
New Noise Metric

2) H.265 quality indicators
New SRC and subjective scores

3) Crowdsourcing experiment:
Objective: set alarm threshold for blur, blockiness, contrast, and noise.
Tested on images

4) Go to: <http://vq.kt.agh.edu.pl/> for this nice software that implements these metrics. The crowdsourcing software can also be made available.

Q&A:

Have the indicators been tested on compressed material or uncompressed
- On both compressed and artificial blur for example
- For crowdsourced experiment, solely through artificial blur. Using compression, distortion would be combined and not separated out.

It is suggested to down and upsample instead of introducing artificial blur.

Would the sharp text next to the picture influence the scoring behavior?
- This was not considered as being a high influence

Did you notice training behavior after a while?
- This has not been analyzed extensively.

Did you measure the time people used to vote?
-Timing information has been captured, but not analyzed.

Is subj. test available?
- It can be made available if there is interest.

Is GitHub an option?
- Executables are downloadable from the website.

**QART**March 2nd 2:25PM

Mikołaj Leszczuk (AGH) & Lucjan Janowski (AGH): Revision of P.912 [file=VQEG\_QART\_2016\_111\_P.912\_Revision]

ITU-T Recommendation P.912
Title: “Subjective Video Quality Assessment Methods for Recognition Tasks”
Published: 2008

SG9 Meeting, 21-28 Jan, Geneva:
- Clause 7.5 (“*Crowdsourcing Environment*”): Whole process described from software preparation to running experiment.
Crowdsourcing experiments are advised to be kept short.
2000$ delivered 40 answers per sequence (960 sequences)
Randomization of answer buttons was investigated to randomize votes on guesses of the participant.
There were no complaints on randomizing positions, and there was no significant difference between the random button order and the fixed button order.
Resolution CIF-> VGA changes prob. 64%
lighting: sun light 38 times better results
- Clause 8.1 (“Data Analysis”): Subject Screening, Recognition Probability as a Function, Comparing Different Conditions
- Seeking for final consent (approving, closing)

Q&A:

In the stars test for crowdsourcing, it is suggested to also use the inverted test, like white stars on a white background.

What is overall opinion on recognition procedure?
- Only 7 objects enables people to easily “recognize” things
- Language is a barrier
- people memorize objects well
- creating realistic cases is difficult; plate recognition is better in that sense.

Do you think changes like 7 to 40 objects fixes the recognition tests?
- Characters are a way to get a lot more than 40 objects, and by using combinations, end up with millions of combinations. The problem with cars was that owner needed to provide written consent.

Would crowdsourcing the images help?
- Indeed, creative commons would be a possible solution.

**VIME**

3:45PM

Mikołaj Leszczuk (AGH): Automatic labeling of images from the VIME Flickr Group [file=VQEG\_VIME\_2016\_109\_Automatic\_Image\_Labeling]

The system developed in collaboration with AGH (IMCOP) has been connected to the VIME Flickr database, in order to automatically label the pictures.

* Face detection and head count
* Planned: Talking head detection
* Possibly not relevant for VIME: Face recognition
* Text detection
* Possibly not relevant for VIME: Logo detection
* Possibly not relevant for VIME: Watermark detection
* Color of clothes detection
* Near duplicates detection/ even sub pictures
* Bokeh effect detector
* Face orientation detection
* Smile detection
* Unshaved faces detection
* Red eye detection
* Nudity identification
* Also the MOAVI indicators
* Dominant colors
* Landmark object recognition
* Skyline detection
* Emotion detection
* Planned: Age classifier, indoor/outdoor, day/night

Q&A:

How to interact with IMCOP?
- the interaction can be discussed, but at this moment , the flickr group is crawled from time to time and pictures are automatically tagged by IMCOP.
- Currently the tagging has not taken place yet. Planned by the end of June

What about the detection performance?
- Performance figures are between 70% for the worst algorithms and 90% accuracy for the best ones.

What is the size of the database on which these indicators are tested?
- between 100 and hundreds of pictures.

## Thursday

|  |
| --- |
|  |
|  |
| **VQEG 03.03.2016** |

* Zhenzzhong Chen and Phil Corriveau will be co-chairs on "immersive media group"; we will look for a third co-chair
* eLetter :
	+ VIME eLetter recently published
	+ Agreement on new topics:
		- Immersive Media
		- Visually lossless compression
	+ Group co-chairs and eLetter editors will coordinate this effort
	+ Further discussion by e-mail
* Presentations :
	+ #9: Kjell Brunnström (Acreo), AVHD, Objective characterization of adaptive bitrate videos of long duration and the relation to the subjective quality of shorter adaptation events, VQEG\_AVHD\_2016\_116\_Adaptive\_Streaming\_Characterization - Q&A:
		- Q: Asking for further explanations on confidence intervals?
		- A: Explanation provided.
		- Q: What do subjects think on the study?
		- A: They love it.
		- Q: What is distribution of scores for individual PVS?
		- A: Available in the journal paper and Samira Tavakoli’s thesis.
			* §  S. Tavakoli, "Subjective QoE Analysis of HTTP Adaptive Streaming Applications," Ph.D., Universidad Politecnica de Madrid, Madrid, Spain, 2015.
			* §  S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. Garcia, "Quality of Experience of Adaptive Video Streaming: Investigation in Service Parameters and Subjective Quality Assessment Methodology," Signal Processing: Image Communication, 2015.
		- Q: Is dataset going to be available?
		- A: Subjective data can be shared; ask Kjell about videos.
		- Q: Is display order affecting the scores?
		- A: Yes, last PVS correlate much worse.
		- Q: Would introducing pair-comparison influence the results?
		- A: Discussion.
		- Q: Do switching events influence the quality perception?
		- A: Some results confirm this.
	+ #10: Kjell Brunnström (Acreo), AVHD, Applicability of existing objective metrics of perceptual quality for adaptive video streaming, VQEG\_AVHD\_2016\_117\_Adaptive\_Streaming\_Objective\_Metrics - Q&A:
		- Q: What parameters should be used for subjective tests?
		- A: The test design parameters were discussed with Swedish TV broadcasters (SVT and TV4) to be inline with parameters they used for their streaming services.
		- Q: Could the correlation within some windows be better?
		- A: To be checked.
		- Q: Perhaps low performance is caused but not taking into account behavioral aspects.
		- A: Yes, this is possible. Some of the metrics are completely out of scope and not trained for this case.
	+ #11: Lucjan Janowski (AGH) & Margaret Pinson (NTIA/ITS), AVHD, Project update: experiment designs to avoid scene reuse, VQEG\_AVHD\_2016\_118\_Novel\_Experiment\_Design\_Update:
		- Presentation:
			* Problem Statement
			* Results
			* Summary
		- Q&A:
			* Q: What are the bitrates of the codecs?
			* A: Explained actual values.
			* Q: What software was used for compression?
			* A: It does not matter.
			* Q: Does „better” mean filling better the chart surface?
			* A: Not clear.
			* Q: How would the surface change if contained to the lower and the better ones?
			* A: It is not easy to get a good representation of the entire space. it only guarantees there will be some lines inside the space.
			* Q: What is the motivation to use same scenes?
			* A: It is just it was done for years.
			* Q: What was the question?
			* A: Ignore scene topic.
			* Q: Reusing the same scenes may cause to amplify rare artifacts?
			* A: Yes, good arguments.
			* Q: Is it like for some cases we should use traditional experimental design, and for other cases - the new ones?
			* A: Yes, this is basically the conclusion.
* AVHD-AS validation project:
	+ Changes proposed in an Introduction (change tracking), no objections observed.
	+ Sentence added on duration of the sequences (1..5 minutes), no objections observed.
	+ Changes (in change tracking) in „2.5 Model Input” on additional input, discussion happened, but finally no objections observed.
	+ Changes (two paragraphs added in change tracking) on viewing distances, no objections observed.
	+ New section added („2.8 AVHD-AS Data Analysis”, in change tracking), discussion happened, but finally no objections observed.
	+ Changes in (or, actually, contributions to) „2.13 Schedule” milestones (in change tracking), discussion with Chulhee Lee happened on liaisons. **Action point**: VQEG should send a liaison to other standardization bodies (a week after the meeting) on progress on AVHD-AS. Furthermore, discussion happened with Chulhee Lee, Silvio Borer and Mathieu Carnec, causing some minor changes in the in the schedule originally proposed by Christian Schmidmer (noted in the AVHD document). No further objections observed.
	+ Changes in „2.14 AVHD-AS Fees” (marked in change tracking) - this topic is planned to become a private discussion, no objections observed.
	+ Discussion on revisiting AVHD in the afternoon in some time is saved. No objections observed.
* RICE:
	+ Presentation by Kjell Brunnström:
		- Name change proposal to PsyPhyQA (Psycho-Physiological Quality Assessment). Agreed.
		- Proposal to add Naeem Ramzan as the co-chair. Agreed.
	+ Presentation by Naeem Ramzan:
		- Q: Is there a link between emotions and the quality?
		- A: That needs to be explored as well.

(started note-taking while Naeem was presenting, 2:00pm)

Elaine Jin asked why she couldn’t see the relation between emotional status and observations/detections. Naeem commented that annoyance can be observed, but not necessary emotion.

Patrick mentioned that there is a publication on ICIP 2014 on the same issue that shows that relationship between emotion and QoE

\* RICE / PsyPhyQA - Sebastian Bosse presented “Using Steady-State Visual Evoked Potentials for Electroencephalography-Based Image Quality Assessment”

First part of the presentation was based on the SPIE and ICIP prior work

Second part of the presentation was an update that hasn’t been published yet

During the presentation:

- Michelle Saad asked whether the signal presented was from 1 or more subjects. Answer was that the plot was for 1 subject, but the prediction was based on average

- Michelle Saad asked whether the position of electrode placement near the visual cortex is correct, but could instead be that visual cognition takes place in another part, such as the frontal lobe. Answer is that there is sufficient anti-correlation to support the current hypothesis

Questions/comments were pushed to private discussions

\* AVHD - Lucjan presented “The Accuracy of Subjects in a Quality Experiment: A Theoretical Subject Model”.

During the presentation:

- Patrick asked how do you define the “true quality of an experiment” and Lucjan responded that this is the limiting average value when number of subjects grows to infinity

- Patrick asked when there were repetitions, whether these were randomization and Lucjan and Margaret said there was a total of 2 randomizations.

After the presentation:

- Margaret commented this work can be used to select how many subjects can be used to do the visually lossless experiments

- Ludovic asked whether normalizing each subject’s distribution would produce similar or different results; Lucjan responded that user normalization is more of a measuring tool calibration process instead of subject bias removal.

- Ludovic asked whether the range (1-5, or 1-9) would affect the study; Lucjan commented that this wouldn’t make a significant difference

\* VIME - Margaret presented “Image Quality for Public Safety”

During the presentation:

- Jorge asked whether there is an objective model to obtain a score that after some thresholding one can obtain the acceptability of quality, similar to gesture recognition tasks; Margaret suggested an offline detailed discussion.

- Jorge argues about a machine learning approach that through training can provide a result & confidence interval. Alternatively, you can do statistical analysis. Margaret commented that this is a partial solution.

After the presentation:

- James asked whether the forensics audience has any standards they work with (such British standard BS 6276 for surveillance video). Margaret commented that there are some associations (like LEVA) offering classes and certifications, but still to determine if there are standards; at best such standards would be limited to the US. There is also prior work done at NTIA through Video Quality In Public Safety. Work seems to be suspended lately. Moving to video analytics from video quality.

- James suggested to show to interested groups what is already available through VQEG and receive feedback.

Lucian - New journal: QUALITY AND USER EXPERIENCE by Springer

\* VIME - Elaine Jin presented “Overview of CPIQ” (IEEE P1858)

Targeting camera design and test engineers.

Compete with DXL-Mark score.

Auto Exposure (AE metric - EI16) example presented

Handbook of image quality by Eastman-Kodak was adopted for quality rating (SQS = standard quality scale, representing JND from 32 - 0).

The combination of 7 features is through JND quality degradation and then multivariate formula that penalizes worst performance. Final outcome is a global JND number, that can be translated to a star-rating scale (e.g. 0-5)

During the presentation:

- Question from Roberto Fonseca asked whether video is included, Elaine responded it will be added in version 2

- Question from Chris was about how to use the quality ruler, and Elaine responded they did, by using a sharpness scale to compare against exposure or other features, rather accurately

- Lucjan asked how to check the performance of the quality ruler method, and Elaine said that they relied on averages to obtain the results

- Alexis asked whether de-bayering was performed before/after raw image capture and Elaine responded that a dedicated MATLAB toolchain was used to do de-bayering and obtain results.

- James asked what are the different JND figures reported per component; Elaine responded that it is the result of different performance w.r.t. the different criteria.

- Lark asked whether there is any masking effect in the creation of the score; Elaine responded that masking is performed by the multivariate formula itself.

After the presentation:

- Steve asked about video quality ruler; answer - Judith Redi presented this during last VQEG meeting

- Jorge asked about the spatial resolution metrics and whether it depends on sensor resolution and technology, and Elaine responded it relates to perceived sharpness.

\* UHD - Glenn presented “Subjective Evaluation of 4K-resolution video content”

During the presentation:

- Alexis asked whether the 63% of people who preferred 4K as sharper from H means the others said “same” or they preferred HD; Glenn said that the rest chose HD.

- James asked whether there was repetition; Glenn confirmed they saw the content twice

- Roberto Fonseca asked whether the time mix of 30, 50, 60 fps content makes a difference; Glenn said he had no answer

- A number of people asked about the quality of the sources - Glenn explained some were film, others RED-camera

- Ioannis asked about color space; Glenn responded YUV420, for convenience

- Alexis asked about filters used; Glenn responded Lanczos

- James asked whether subjects were screened for visual acuity; Glenn responded positively

- Alexis asked about randomization; Glenn responded positively and explained the methodology

After the presentation:

- James asked about 8K and whether the methodology is valid; answer was positively, but that in order to be able to tell the difference you need appropriate kind of content

- Alexis asked about the quality of the display; Glenn responded they were calibrated

\* UHD - Patrick presented “Added value of UHD over HD”

During the presentation:

- Naeem asked about viewing distance; Patrick responded 1.5H for UHD and 3.0H for HD

- James asked what would be the gaze fixation if using 480p displayed on SD, HD and UHD-resolution displays; Patrick responded it would be different in these cases. Viewing attention depends on viewing field.

\* UHD - Florence presented “Demo on UHD content from Sky (sports & drama) for research purposes”

VANGUARD VIDEO has **Visual comparison tool** (free 90-day trial) that allows stitching two different videos side-by-side with a ruler

During the presentation:

- Kjell asked whether audio is available; answer is no

- James asked about rights; answer is through CDVL

Meeting was adjourned at 5:45pm

## Friday

Meeting opened 9:10am

\* AVHD (Chris) - Discussion on AVHD-AS (see document: AVHD-AS Model Requirements)

5 Proponents so far specified interest in participation:

* Opticom
* SwissQual
* AccepTV
* Cardiff Metropolitan University
* Stream Owl

Kingston University and Yonsei University are potential proponents

Editing of the document, with input from attendees and discussion took place. See the updated document in meeting files for changes and the document as agreed.

Shahid led a discussion on the subjective testing procedure.

ACR testing with related sequences is the primary method proposed.

**Agreement** to use ACR (without hidden reference removal), and the related sequence experiment design as much as possible.

**Agreement** 1/3/5 min databases, no mix of durations, in a single experiment

**Agreement** P.913 testing standard, not for viewing distance, for which proposal is 3-5H (living room experience) for TV testing -> agreed, with viewing distance TBD.

Coffee break

\* VIME (Michelle)

- VIME image database on Flickr (read e-letter for details on how to upload images)

- Mikolaj is creating a crawler to process all images on VIME to analyze and calculate features for each image; this data will be uploaded to Flickr. This process can be repeated when new images are uploaded. TBD whether this will be a manual or an automatic process.

Another item is to decide what other algorithms need to be integrated into the VIME analysis software.

Discussion took place on how to handle ground truth data.

- New questions were raised on supporting

(a) Panoramic images

(b) HDR images

(c) Selfies

Mikolaj mentioned prior experience on panoramic images.

Phil commented on selfies, panoramic and HDR as new application areas of value. Selfies possess challenges (privacy); the other two should be put on the table.

Margaret agreed with Phil, adding that contributions to ITU-T would be useful.

Ioannis questioned whether panoramic images need special treatment. Michelle mentioned that the stitching process makes a difference and that would be the main feature to be studied.

Lark asked on the specifics of the stitching and display characteristics; Michelle commented that this will be part of the study.

Jorge suggested to identify the new experience that one targets and then develop subjective test methodologies and metrics.

Elaine commented that stitching is not the only factor, but 3A (AF, AE, AWB) are also very important factors to study.

Margaret mentioned there is already 1 panoramic image on the VIME database.

Asked the group about interest in these 2 new features - most expressed interest on HDR.

Discussion on the availability and type of consumer HDR pictures took place.

Jorge commented that panoramic images can be studied in terms of saliency and stitching accuracy.

The issues will be further discussed in VIME conference calls.

\* AVHD (Shahid) - AVHD-AS Subjective testing methodology

Draft document was presented and discussed. Edits are included in the updated document, available in the meeting files.

Meeting was adjourned 12:30 pm.