 VQEG Meeting Minutes - Stockholm July 2014

# Monday AM, July 7

## Group Updates

### more

with some text

# TUesday AM, July 8

Short discussion about the possible next meeting. INTEL has offered to host in California. Suggestion to put it together with the SPIE Electronic Imaging, but this year this was a problem with people not being able to stay away too long. Together with VPQM this could be almost 3 weeks. VPQM is not decided yet.

## QART

### Overview of QART

Mikolaj gives an overview of the QART project, see presentation. QART targeting Task Recognition Video (TRV). This is partially standardized in P.912, but this is not sufficient based on the results in the project.

Can you force people to describe their experiment? Could that be put into the recommendation? Maybe not limit it to licence plate recognition. Description of the particular task at hand is more important. Ask just one thing at the time and simple questions. Also take into account the training of the subjects. Also study how professionals are using the videos in practise e.g. doctors. In sign language studies are limited to a small subset and then extrapolate to a larger set. This may not be appropriate. The standards say that entire ranger. Does that means 6 sigma or 9 sigma or. This should possibly be clarified.

Non clear paragraph should be clarified. Is the camera parameters specified. That may be necessary to clarified. Action point to call for input and get feedback from expert eg Qualcomm

Impact of position of choice button. This is not specified in the standard. Upper right corner was used more often when people of uncertain. But it may also have been the selection of the alternative in the list. If the alternatives are not equally likely the experiment is biased.

Comparison scale with “about the same” is abused even if people can identify. Should warn against “unsure”. Leaving out “unsure” may give you more accurate data, but you need equal probability of alternative. A solution could be to introduce a confidence scale, but psychologist are using forced-choice so that would most likely be the best route. Pre-test could confirm the distribution.

Possibility to use Fuzzy logic. Specify the task first then the analysis can be specified.

Give input to QART if you are not happy about P.912.

People are not motivated. What is the solution. Idea gamification. Payment is one solution and motivating the subjects another.

Suggestion to update P.912 to refer to P.913 instead of P.910.

Statistical analysis should be specified in the Recommendation. Maybe also put in post-screening rules.

A more detailed document has been submitted to VQEG. The plan is to submit to ITU meeting in September 2014. Try to reach out to those who use the P.912 in advance to get them involved in the process, which may help not giving problem in the future.

## AVHD

How should AVHD proceed in the audio-visual part? Collaborative is suggested as the best. Reference best practise is also suggested.

**Decision:** will go for collaborative approach

How should HD proceed? Most are working on adaptive streaming among proponent, so what is the real interest in a pure HD test. Is there any interested in a quick HD test. The idea is to go for UHD

**Decision:** Tabled validating HD video only models

HD method for new codecs (VP, HEVC..) may be validated within the UHD project

Going through the test plan of audio-visual testing.

Editing is done directly in the test plan. No decisions is taken at this point, rather the full document will be approved when it has reached a stable state later this meeting.

# TUESday PM, July 7

## Something else

### Presentation from Swedish ict – acreo representative

Introduction on Swedish ICT and the environment where the current meeting is hosted. Anders Berntson.

## JEG

### Logistics

Several persons online:

* Amy Reibman
* Enrico Masala
* Kongfeng Zhu
* Naeem
* Rahul Gaurav

2 Remote presentations planned.

### Introduction (by Marcus Barkowsky)

Introduction on JEG was given:

* similar structure as H.261 🡪 H.265 community.
* Open database

Ongoing work is layed out in the introductory presentation

Q: Dataset: 10 source sequences. Reason not to use existing sources from HD project that has 50 sources? 10 sources is limited for doing subjective tests.

A: work in progress. Now 31 HD sequences (downscaled from UltraHD sources) are being processed. As soon as processing power is available, the 50 sources referred to by Chulhee wil be added.

Q: Why not limit number of PVSes and have more sources instead.

A: We are working on more SRC, but need extended number of PVSes/HRC as well

### No-reference video quality assessment based on artefact measurement and statistical analysis (by Kongfeng Zhu)

Proposed model:

* Kernel-based analysis of images.
* Feature extraction: kurtosis, smoothness, sharpness, historgram noise, Mean Jenson Shannon divergence (MJSD), Blockiness
* Predict video quality: 2 steps
  + Temporal pooling: get 6 features
  + Infeed 6 features into neural network: 6 inputs with 20 hidden nodes, 1 output. (fully connected network)

Used model on 4 databases (selection in slides)

Neural network training:

* Cross-validation strategies based on size of database (k-fold vs. leave-p-fold-out)
* Training models: linear model, support vector machine, multilayer perceptron

Q: (James): colorspace of analysis?

A: luma only in this presentation, included color in other work which has not been published yet.

Q’s: how were the training set and validation set selected.

A: the \*fold\*-approaches were repeated with new selections of training set among the total database.

### Machine learning for quality assessment (By Adriaan Barri)

Subtitle: adopt – adapt – improve???

Parameters: Audience, device, content 🡪 optimize quality measures: choose quality indicators + weights

Proposal: use machine learning within the ‘quality network’

Linear regression: too rigid. Weights must adapt to the input. 🡪 neural network with LAF (Locally Adaptive Fusion): fusion with certain targets (indicator values/local optimization). 🡪 then select the best fitted model for the indicator values for the current indicator area you are for the image.

Technical information: [www.locally-adaptive-fusion.com](http://www.locally-adaptive-fusion.com)

How to avoid overfitting?

Q: code we can run?

A: on request you will get the source code..

Q: overfitting issue solved?

A: not solved, but you can easily spot it when it applies to your trained system. (2 dimensional plot: wigly = overfitted, instead of 80 dimensional plot)

### No-reference consumer-oriented image/video quality assessment: from theory to a standard (by Michele saad and Philip corriveau)

CPIQ: Camera Phone Image Quality 🡪 IEEE standards working group. Comparison between camera qualities from (mobile) devices.

Q: Laboratory approach is less good than customer approach? Are other customers more to be trusted than 3th party laboratory

A: Expensiveness of setting up a lab + full reference approach is not perfect too.

Q: Integrator point of view: embedders of cameras on mobile phone

A: DXO – Imatest, AMD, blackberry, huawai, MIT…. and the whole list of about 20 big players.

Q: do we need a standard?

A: objective modelling approach currently: create test plans and labs, industry tailors to model. Now more consumer oriented resulting in a larger database of images that is relatively easy/cheaply obtained.

Other comments: Blurred sections overlay, customer profile dependency

Decision: this topic will be a subject under JEG—a new subgroup:

* unanimously,
* indicated co-chairs: Michele Saad, James Goel, Quan Huynh-Thu (expressed interest --subject to approval by organisation)
* acronym needs to be generated as well as a description of the topic

### HEVC (By glenn van wallendael, nicolas staelens)

Update on current PVS generation.

Components can be submitted, integration will be looked at when components become available. (Creating an API is a subject for further work)

### Simulation of robust hevc decoding (By Enrico masala)

Simulate packet loss in the decoder 🡪 avoid expanding the database exponentially by adding one additional impairment

Concept is found very useful by multiple people in the meeting, as it solves a lot of issues:

* Crashing decoders
* Management of difficult types of

### First analysis on the large dataset what we can learn by << only >> comparing objective measurements (by enrico masala)

Very interesting analysis. See presentation: ???

### Equal quality threshold analysis (By Marcus Barkowsky)

*Preliminary results* were presented ???

Threshold-based comparison of 2 objective metrics. Comparing the false ties (one metric says A = B, other says A < B or B < A; weak error) to false orderings (one metric says A > B, other says B > A; strong error).