VIDEO QUALITY EXPERTS GROUP

Progress report 2014 v.1

For more information contact:
Kjell Brunnström kjell.brunnstrom@acreo.se Co-Chair VQEG
Margaret Pinson margaret@its.bldrdoc.gov Co-Chair VQEG
Arthur Webster webster@its.bldrdoc.gov Co-Chair VQEG

Introduction

Very good progress was achieved during 2014. One major competitive project was finalized, leading to a series of new recommendations within the ITU. The management team of VQEG was strengthened by including Margaret Pinson, NTIA/ITS, as a new and additional Co-chair of VQEG.

The 2nd eLetter was released focusing on verification and validation, which has been an important topic for VQEG from its start.

Two regular meetings were held during 2014. One was co-hosted by NTIA/ITS and Rohde & Schwarz in Boulder, Colorado, USA, January 21 - 24. There were 35 participants, including remote participants. The other meeting 2014 was hosted by Acreo Swedish ICT AB in Stockholm, Sweden, July 7-11. A total of 44 people participated in this meeting, either physically or remotely. Furthermore, an ad-hoc 3-day meeting was arranged by the 3DTV project and hosted by the University of Nantes in Nantes, France, December 8-10, 2014.

In 2014, VQEG work emphasized collaboration to develop new and improved methods for measuring video quality. Meetings emphasized presentations that described research underway or newly finished. These efforts are divided between objective metrics (JEG-Hybrid, MOAVI, and VIME) and subjective methods (3DTV, AVHD, QART, RICE, and UltraHD). Research on 3DTV subjective methods is mature enough that VQEG's knowledge is being passed to ITU-T Study Group 9 for proposed inclusion into new Recommendations.

VQEG has 11 active workgroups and 4 support groups. Their activities and progress during 2014 will be described below in this report.

Active Workgroups

3DTV (3DTV)

Co-chairs: Marcus Barkowsky (IRCCyN), Patrick LeCallet (IRCCyN), Quan Huynh-Thu (CiSRA)

Goals:

The 3DTV Project investigates how to assess 3DTV subjective video quality. The 3DTV workgroup has established 4 work items, which are progressing in parallel:

1. Define suitable methodologies for subjective quality assessment of stereoscopic 3D video:
   A unique set of video sequences is being produced to conduct the experiments (GroTruQoE dataset). The first step is to conduct a large-scale experiment using the pair-comparison methodology, as we are confident that subjects can provide easily a judgment of overall preference. The second step will use the results of the pair-comparison testing as a groundtruth database to investigate which more time-efficient subjective testing methodology can be used to predict the results of the pair-comparison test.

2. Investigate the influence of viewing environment, test set-up and display equipment on subjective quality:
   The purpose is to investigate restrictions which should be imposed on the subjective testing environment and viewing conditions to obtain reliable results across test labs. A unique set of 3D test sequences (COSPAD dataset) was used by different research labs. Each test lab studies variables related to viewing environment/display equipment, and report experimental results for comparison with other labs' results
3. Objective video quality metrics for stereoscopic 3D:
   3D quality of experience is related to several perceptual dimensions such as visual quality, depth rendering (depth quality, depth quantity), and visual comfort. This first phase of evaluation will focus on the visual picture quality dimension and evaluate full-reference media-layer metrics (use of decoded pixel information)
4. Analysis of frame compatible 3D video format representations:
   In close collaboration with the DVB consortium, VQEG compares the visual impact of Side-by-Side, Top-Bottom, and Tile format in terms of Quality of Experience for Full-HD television transmissions at different bitrates. Various subjective assessment methodologies and appropriate statistical analysis tools are discussed.

Achievements 2014

- A dedicated workgroup meeting December 8-10 was held in Nantes to discuss the DVB test plan and contributions to three ITU standards:
  1. P.3D-sam: Subjective Assessment Methods for 3D Video Quality
  2. P.3D-disp-req: Display Requirements for 3D Video Quality Assessment
  3. P.3D-fatigue: Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video
- The DVB test plan was established and goals were clarified in regular telephone conference calls
- In the Ground Truth effort GroTruQ3D1, we have established the PVS and the first phase of subjective assessment with Paired Comparison is ongoing, collecting about half of the votes

Plans for 2015

- Conducting and analyzing the DVB experiment
- Analyzing results of the first phase of the GroTruQ3D1 data and conducting the second phase
- Continuing on the ITU standardization documents for the three 3DTV standards with the final draft version available during the VQEG meeting in August/September/October, including the results of the GroTruQ3D1 evaluation and the COSPAD evaluations

Audio-Visual HD Quality (AVHD)

Co-chairs: Chris Schmidmer (Opticom), Quan Huynh-Thu (CiSRA), Margaret Pinson (NTIA/ITS)

Goals:

The AVHD Quality workgroup has 3 active projects:

- Video-only quality in HD resolution
- Audio-visual quality of HD videos with accompanying sound
- Adaptive video streaming quality assessment methods

Achievements 2014

- Video-only quality in HD resolution:
  - This project initially started as the direct successor of the HDTV Phase I project
  - It was decided to put this effort on hold for the time being, as most participants are focusing more on the audio-visual quality and adaptive streaming projects. This project may be merged in the future with the UHDTV project as interest towards ultra-HD is growing and new codec such as HEVC may be evaluated directly on UHD.
- Audio-visual quality:
  - Several presentations reported results of new experiments investigating new audio-visual quality assessment methodologies
  - It was decided to focus on the problem of integration of audio and video qualities for the prediction of overall audio-video quality
  - Survey of studies was presented on integration of audio quality and video quality into an overall audio-video quality score. Many different integration functions have been proposed in the literature, which may highlight the strong effect of content-dependency on the integration between the two modalities
  - Software package was made available to help research efforts: the Web-Enabled Subjective Test (WEST) software package offers a solution to the problem of gathering subjective testing data from multiple locations and multiple portable or computing devices. The software is freely available for download from the NTIA/ITS website.
- Adaptive video streaming quality:
  - This project started in 2014. Discussions have targeted the definition of the scope of work, in particular how this project can complement the P.NATS work in ITU-T SG12. Interest is going towards the quality evaluation of very long sequences (e.g. several minutes), which would require investigation of new subjective testing methodologies or refinement/adaptation of existing ones.
  - Several studies on quality of adaptive streaming were conducted by VQEG members and reported. One study compared results obtained in controlled lab test and results obtained from crowdsourcing.
  - A joint VQEG/Qualinet activity was initiated to survey existing literature on the topic, which resulted in a QoMEX paper ”Quality of Experience and HTTP adaptive streaming: a review of subjective studies”

References

1. M.–N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnstrom, and A. Raake, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," 2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX), 18-20 Sept. 2014. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6982310&matchBoolean%3Dtrue%26searchField%3DSearch_All%26queryText%3D%28Quality+of+Experience+and+HTTP+adaptive+streaming%3A+a+review+of+subjective+studies%29
2. A. Catellier and L. Connors, "Web-Enabled Subjective Test (WEST) Research Tools Manual," NTIA Handbook HB-14-501, January 2014. Available: http://www.its.bldrdoc.gov/publications/2748.aspx. Software: http://www.its.bldrdoc.gov/resources/video-quality-research/web-enabled-subjective-test-west.aspx

Joint Effort Group - Hybrid Perceptual Bit-Stream Measurement (JEG-Hybrid)

Co-chairs: Marcus Barkowsky (IRCCyN), Lucjan Janowski (AGH University), Nicolas Staelens (Ghent University-iMinds-MMLab), Glenn Van Wallendael (Ghent University-iMinds-IBCN)

Goals:

The JEG Hybrid Group is an **open collaboration** working together to develop a **robust Hybrid Perceptual/Bit-Stream video quality measurement tool**.

Perceptual video quality measurement has been tackled by many researchers in the past. JEG-Hybrid aims to **unite their independent work into a framework** where individual ideas of modeling the Human Visual System become verifiable and measureable.

Bit-Stream information may be seen as a primary or supplemental information source for video quality assessment. As a primary source, properties such as bit-rate or the quantization parameter may be exploited. As a supplemental source, the video bit-stream delivers information that is otherwise often calculated on the decoded video, notably motion information or frequency analysis. **Rigorous analysis of the value of these information sources** is required. Within JEG-Hybrid bit-stream information is easily exploited beyond its current usage in parametric models and bit-stream models by using a simple interface for interpreting its data in XML files.

Hybrid approaches combine various indicators. These indicators may be quality indicators, degradation indicators, content indicators, and so on. Challenges include the **frame-exact synchronization of bit-stream information with perceptual information** obtained from the decoded video, or innovative approaches on **stable combinations of a large number of indicators** on a comparably small training set in machine learning algorithms.

Robustness of video quality measurement stems from verification and validation on a large dataset. The JEG-Hybrid team maintains and continuously improves a **large-scale database of encoded video sequences** with recent video coding algorithms. Innovative research is performed to **verify prediction performance** of objective measurements with this database because subjective assessment can only be performed on a comparably small and well-chosen subset.

Achievements 2014

The JEG-Hybrid group has advanced on:

- Creation of a packet loss robust reference decoder version for HEVC (HM)
- Review of recent objective video quality measurements [2]
- Creation of large scale database
    - About 60.000 HEVC coded video sequence database
    - Annotation of the AVC and HEVC database with objective measurements (PSNR, SSIM, VQM, PEVQ)
    - First analysis of the objective quality indicators published [3]
    - Future research topics definition [1]
    - Extension of the existing database with packet loss scenarios

- First version of an open source reliable video player for video quality assessment (VQEGplayer)
- An HEVC reference decoder capable of writing bitstream information to an easy to read XML file (HEVC-TO-XML converter)

Plans for 2015

- Advance on general goals:
    - Collect and implement FR, RR, and NR measurement algorithms, reach out to individual model researchers
    - Extend the set of Bitstream-Based and Pixel-Based features and indicators
    - Continue the research on adapted machine-learning algorithms
-
- Potentially organize a research and implementation event (ex. Isolated work-group meeting of JEG-Hybrid)
-
- Perform subjective assessment(s) on a subset
- The database will be extended with a larger number of contents (SRC) that will be processed in the same way as the existing SRC to improve on the general applicability of the large-scale database
- Broaden the scope towards DASH applications, i.e. add tools and HRC specific for DASH
- Improve on the reproducibility by including all HRC into the VirtualBox
- Integrate the objective measurements and objective indicators into a MySQL or NoSQL database

References

1. Barkowsky, M., Sedano, I., Brunnström, K., Leszczuk, M., & Staelens, N. (2014). Hybrid video quality prediction: reviewing video quality measurement for widening application scope. Multimedia Tools and Applications, , 1–21.
2. Shahid, M., Rossholm, A., Lövström, B., & Zepernick, H. - J. (2014). No-reference image and video quality assessment: a classification and review of recent approaches. EURASIP Journal on Image and Video Processing, 2014(1), 40.
3. Van Wallendael, G., Staelens, N., Masala, E., Janowski, L., Berger, K., & Barkowsky, M. (2014). Dreamed about training, verifying and validating your QoE model on a million videos? VQEG eLetter, 1(2), 19–29.
4. Barkowsky, M., Masala, E., Van Wallendael, G., Brunnström, K., Staelens, N., & Le Callet, P. (2015). Objective Video Quality Assessment — Towards Large Scale Video Database Enhanced Model Development. IEICE Transactions on Communications, E-98b(1), 2–11.

HDR (High Dynamic Range)

Patrick LeCallet (IRCCyN), Elaine Jin and Philip Corriveau (Intel Corp)

Goals

Human eyes can perceive approximately 5 orders of luminance magnitude at a time, but current capture and display technologies are limited to at most 3 orders of magnitude.

The aim of HDR imaging: capture and display these 5 orders of magnitude with high fidelity.

The aim of the VQEG HDR project is to develop methods for assessing the quality of HDR video.

HDR technologies aim at filling the gap between capture and display technologies and the abilities of the human visual system. This will provide a more realistic visual experience compared to current Low Dynamic Range (LDR) imaging. All of the current technology on the market is LDR but it is felt that within the next few years HDR will be ready for commercial release and we need to be ahead of it to understand the user benefit and how to measure quality.

Achievements 2014

There were 4 main accomplishments completed in 2014 and the plans for 2015 are coming in line well.

First was that the group created a database that has enabled the testing for how tone mapping affects human attention and modification of artistic intent. This work enabled a comprehensive evaluation of objective methods and solid statistical analysis was completed. The database is available for other researchers or efforts.

Secondly another dataset was created for the investigation of tone mapping operators for quality. This yielded two results around the use for investigating quality issues in local and global HDR codec optimization. Also it enabled the creation of a database with corresponding subjective scores. It is still up for consideration on if the database will be able to made available for other researchers or efforts.

Third there was a drive in the core metrics development for HDR. Tools were created for objective HDR assessment validation of a new version of the tool HDR-VDP-2.2 which has been released for use in the industry.

Lastly the group completed testing in support of validation of an HDR metric extended to Video. All previous tools and metrics have been for still images. HDR-VQM has been extended to enable the algorithms scale to motion video.

Plans for 2015

HDR image capture project deliverables

- Create a sharable data base with videos recorded with HDR. Work with legal to get it available to wide use within VQEG and the Industry.
- Survey existing methods for HDR image capture
    - Still and video

    - Market segments
- List key attributes that may impact image quality for each of the methods listed above
- Identify key performance indicators (KPIs) for the list of key attributes

Hybrid Perceptual/Bitstream project

Co-chairs: Silvio Borer (SwissQual), Chulhee Lee (Yonsei University)

Goals:

The goal of the hybrid project is to evaluate hybrid models (FR, RR, NR), which use bitstream data and the decoded video signal as input.

Achievements 2014

Four proponents submitted models. The validation process was finished in mid-2014. The Hybrid Final Report [1] containing details about the evaluation method and model performance was written and approved. The hybrid models proposed for evaluation are now part of the ITU-T recommendations J.343, J.343.1 – J.343.6 [2].

Plans for 2015

At the moment, there is no planned continuation of the project.

References

[1] M.Pinson, S.Borer editors (2014). *Hybrid Perceptual/Bitstream Validation Test Final Report*. ftp://vqeg.its.bldrdoc.gov/Documents/Projects/hybrid/VQEG_Hybrid_Final_Report%20version%202p1.pdf

[2] ITU-T Recommendations J.343, J.343.1-J.343.6, http://www.itu.int/rec/T-REC-J/en.

Monitoring of Audio Visual Quality by Key Indicators (MOAVI)

Co-chairs: Silvio Borer (SwissQual) and Mikołaj Leszczuk (AGH)

Goals:

The MOAVI group is an open collaborative for developing No-Reference models for monitoring audio-visual service quality. The goal is to develop a set of key indicators (e.g. blocking effects, blurring effects, freeze/jerkiness effects, ghosting effects, slice video stripe errors, aspect ratio problems, field order problems or photosensitive epilepsy flashing effects, silence, clipping) describing service quality in general and to select subsets for each potential application. Therefore, the MOAVI project concentrates on models based on key indicators contrary to models predicting overall quality.

Achievements 2014

A new VQEG Joint Effort Group (JEG) Wiki article has been prepared [Video_Quality_website]. The article contains information concerning methods of video quality research. Creators added new metrics and results of research on the video quality. It has also an extensive library of artifacts together with examples and scripts, which enable detection of the items.

As for the now, researchers can freely download and play with the following quality indicators (implemented as MATLAB functions): blockiness, blur, exposure time distortion, interlace, noise, framing, spatial activity, temporal activity, flickering, blackout, pillar-boxing, letter-boxing, brightness, contrast, vignetting, rainbow effect, slicing, aspect ratio, ghosting, ringing, block loss, gray-scale, audio mute as well as audio clipping.

Plans for 2015

Integration of MOAVI indicators within JEG virtual machines has been planned.

References

[Video_Quality_website] VQEG JEG (2014), *Video Quality website*, VQEG JEG Wiki. <http://vqegjeg.intec.ugent.be/wiki/index.php/Video_Quality_website

Quality Assessment for Recognition and Task-based multimedia applications (QART)

Co-chairs: Mikołaj Leszczuk (AGH)

Goals

Users of video to perform tasks require sufficient video quality to recognize the information needed for their application. Therefore, the fundamental measure of video quality in these applications is the success rate of these tasks (such as recognition), which is referred to as visual intelligibility or acuity. One of the major causes of reduction of visual intelligibility is loss of data, through various forms of compression. Additionally, the characteristics of the scene being captured have a direct effect on visual intelligibility and on the performance of a compression operation-specifically, the size of the target of interest, the lighting conditions, and the temporal complexity of the scene. The QART project is performing a series of tests to study the effects and interactions of compression and scene characteristics. An additional goal is to test existing or develop new objective measurements that will predict the results of the subjective tests of visual intelligibility.

Achievements 2014

Recommendation ITU-T P.912 "Subjective Video Quality Assessment Methods for Recognition Tasks", published in 2008, addresses methods for subjective evaluation of Target Recognition Video (TRV). In addition, Recommendation P.912 organizes terminology related to subjective TRV testing, introducing appropriate definitions for the methods of testing (psychophysical experiments).

Unfortunately, Recommendation P.912 is only the first step in the standardization of methods of subjective TRV testing. In the opinion of QART, based on research results (QART's own and independent) and observations conducted during numerous experiments with TRV, many claims of Recommendation P.912 are formulated at too high a level of generality. What's more, selected statements are not supported by research results and are significantly disputable.

The discussion of statements contained in ITU-T Recommendation P.912 shows that some of the findings and observations require the verification of certain provisions of the Recommendation. QART proposed to revise Recommendation P.912 to reflect improved subjective test techniques developed since this Recommendation was approved. Sufficient justification exists to support a new ITU-T work item, and contributions to this topic have been encouraged by ITU-T.

In this situation, QART has taken steps to introduce significant modifications (amendments) to the Recommendation. For this purpose, in order to formalize the procedures, QART has established collaboration with the Polish Ministry of Administration and Digitization. QART has already received a formal nomination as a delegate of the Polish government. The procedure for submitting amendments commenced in 2014 and included source signal, multiple-choice method [b-Green], single answer [b-Leszczuk2014, b-Green] method and subjects [b-Leszczuk2012].

Plans for 2015

Ultimately, the amended recommendations should have a broader scope; expand target testing methods; provide better instruction and training of subjects; improve conditions for testing, statistical

analysis and reporting; and extend the applicability of techniques in the field of crowdsourcing for the subjective assessment of the quality of TRV.

References

[b-Leszczuk2012] Leszczuk, M., Koń, A., Dumke, J., and Janowski, L. (2012), *Redefining the ITU-T recommendation P.912 for requirements of the subjects of quality assessments in recognition tasks*, Proceedings of the 5th International Conference on Multimedia Communications, Services and Security, Krakow, Poland, June, Springer CCIS Vol. 287. <http://link.springer.com/chapter/10.1007%2F978-3-642-30721-8_19>

[b-Leszczuk2014] Leszczuk, M. (2014), *Optimising task-based video quality: a journey from subjective psychophysical experiments to objective quality optimisation*, Multimedia Tools and Applications, United States, January, Springer Vol. 68. <http://link.springer.com/article/10.1007%2Fs11042-012-1161-6>

[b-Green] Green, D. M. (1970), *Application of Detection Theory in Psychophysics*, Proceedings of the IEEE, United States, May, IEEE Vol. 58. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1449656>

RICE (Real-Time Interactive Communications Evaluation)

Co-chairs: Sebastian.Arndt (Technische Universität Berlin), Ulrich Engelke (CSIRO), and Kjell Brunnström (Acreo Swedish ICT AB)

Goals

The Real-Time Interactive Communications Evaluation (RICE) Project is directed towards the development of new methodologies for subjective assessment and objective measurement of interactive communications services. In this effort novel investigation methods based on psycho-physiological measurements are considered in addition to the more common subjective testing methodologies for interactive communication services.

Psychophysiological measurements

Video quality assessment is typically performed using questionnaires, either open-ended or based on psychometric scales, such as n-point Likert scales. As valuable as these studies are, they are based on conscious responses by the participants and often do not provide sufficiently deep insight into underlying perceptual and cognitive processes. In order to gain a deeper understanding about the perceptual and cognitive processes underlying video quality perception, psychophysiological measurements can be performed. For instance,

- Eye gaze tracking provides valuable information about overt visual attention in visual space,
- Electroencephalography (EEG) measurements inform about cognitive activity, such as cognitive load, situational awareness, emotional responses, and covert attention, and
- Galvanic skin response (GSR) provides insight into arousal and hence emotional states.

These psychophysiological responses are not intended to replace well established psychophysical assessment techniques, but to augment them and provide additional sub-conscious information.

- The aim of the RICE project is to establish novel psychophysiology-based techniques and methodologies for video quality assessment and real-time interaction of humans with advanced video communication environments. Specifically, some of the aspects that the project is looking at includevideo quality assessment based on human psychophysiology, including, eye gaze, EEG, EKG, EMG, GSR, etc,
- computational video quality models based on psychophysiological measurements,
- signal processing and machine learning techniques for psychophysiology-based video quality assessment,
- experimental design and methodologies for psychophysiological assessment, and
- correlates of psychophysics and psychophysiology.

Subjective testing methodologies for interactive communication

This part of the project is currently put on hold. New objectives and scope will be decided when this part of the project is restarted again.

Achievements 2014

The RICE project has met a few times online during 2014. At the Stockholm meeting it was decided to have the RICE project also include psychophysiological measurements and put the initial target subjective testing methodologies for interactive communication on hold until a critical mass of interest is available within VQEG.

A plan for a cross lab video quality experiment with EEG was developed during the online meetings and presented at the Santa Clara meeting in February 2015. Tentative interest from the organisations (Netflix) NTIA/ITS and University of West Scotland were expressed after the presentation

The planned study aims on the one hand, at comparing results between labs, and on the other hand, at comparing obtained results between different classes of devices (i.e. low-cost consumer grade EEG products vs high-end clinical grade products). Therefore, typical video sequences are used which have a length of several minutes; while subjects are watching these an EEG will be recorded. In a post processing step the recorded data will be analyzed concerning their power spectral energy using this, assumptions about the cognitive state can be drawn. This inter-lab study aims at validating the methodology of EEG in the context of QoE, which may lead in the future of obtaining non-intrusive correlates of quality.

Several affiliates have expressed their interest in joining the RICE effort on psychophysiological measurement for video quality assessment, including

- Yasuhiro Inazumi, Toyama University, Japan
- Eva Cheng, Royal Melbourne Institute of Technology, Australia
- Ian Burnett, University of Technology Sydney, Australia

Plans for 2015

- Sept 2015: having test plan ready
- Oct 2015: starting subjective tests

Ultra HD

Co-chairs: Vittorio Baroncini (Fondazione Ugo Bordoni), Naeem Ramzan (University of West Scotland)

Goals

Three activities are defined within the scope of Ultra HD project: (1) Creation of Ultra HD database, (2) Defining subjective quality testing methodologies for Ultra HD, (3) Objective video quality metrics for Ultra HD.

1. Creation of Ultra HD database:
   10 4K video contents are available on request. It will be shared through external HDD however the source of the contents is http://medialab.sjtu.edu.cn/web4k/index.html and is free to download.
   Initial testing and subjective evaluation is performed for quality evaluation. H.264/MPEG-4, H.265/HEVC and VP9 codecs are used for quality evaluation.
2. Defining subjective quality testing methodologies for Ultra HD
   Work will be carried out in 2015
3. Objective video quality metrics for Ultra HD
   Work will be carried out in 2015

Achievements 2014

Item 1 above

Plans for 2015

Item 2 and 3 above

VIME (Video and Image Models for consumer content Evaluation)

Co-chairs: Michele Saad (Intel Corp.), Quan Huynh-Thu (CiSRA), James Goel (Qualcomm)

Goals:

No-reference [or blind] image and video quality assessment is still not accurate and robust to be usable in the industry. Current approaches of no-reference image or video quality assessment target the detection and measurement of specific types of degradations. Furthermore, current approaches have focused on the compression and delivery of the content, and not on the capture and rendering of the content. Current approaches also do not consider that image processing can enhance the content, making the quality of the modified image better than the quality of the original image. Increasingly, digital content (images and videos) is being captured by digital cameras and mobile devices, and viewed on a variety of displays and viewing conditions. For these reasons, full-reference approaches are not suited and the requirements for no-reference approaches to measure the quality of images, including capture and rendering aspects, are increasingly important for the industry.

Current databases of images used in research do not contain distortions arising in the capture of images by (compact and high-end) digital cameras and mobile devices (smartphones, tablets). Images and videos captured by these devices contain distortions that are very complex in nature. These distortions may result from the sequence of steps during the capture, from optical system aberrations to the post processing embedded in the devices after the physical light is captured.

The current goals of the VIME project are to investigate and develop:

1) New approaches to subjective study design for the purpose of addressing emerging quality assessment needs (as market and consumer demands evolve).

2) No-reference models for image quality assessment of consumer content with real distortions (as opposed to simulated artificial distortions), including capture and rendering aspects.

3) No-reference models for video quality assessment of consumer content with real distortions, including capture and rendering aspects.

Achievements 2014

VIME was proposed and accepted at the Stockholm meeting in July 2014. In 2014 H2, the co-chairs held a kick-start meeting and then the workgroup had 4 conference calls. Meeting minutes are accessible on the VQEG ftp server: ftp://vqeg.its.bldrdoc.gov/Documents/Projects/vime/

The workgroup has been working towards defining the first goals and targets, as well as identifying the issues to address to start developing a database of content:

- Discussions about the scope of content considered in VIME: decision to exclude medical imaging, [near] infra-red imaging, CGI and artificial patterns.
- Initial discussions on several topics and identified issues:
    - Several options of public image databases were investigated but, so far, none offer a possibility to share freely content among different parties. Freely accessible content is

also not high resolution or high quality. Easiest solution would be for parties to contribute images for the VIME database.
- Technical requirements of the images to be used: format, resolution, minimum level of quality
- Do we need a minimum level of quality? If so, how do we decide on the minimum level of quality of the content to be used?
- Taxonomy of images: how do we categorize the images we use?
- Subjective testing: there is a need to expand the current paradigm of image quality related to low-level processing to more holistic aspect of image quality, including artistic and aesthetic aspects of images. Image quality characterized by the Mean Opinion Score only may not be sufficient and other complementary ways to characterize the perception of the quality of images are necessary.
- Objective quality indicators developed in QART may be useful to characterize the images and for development of no-reference models

Plans for 2015

- Build a dataset of images
- Investigate and report on new subjective testing methodologies

Support Groups

Tools and Subjective Labs Setup (STL)

Co-chairs: Glenn Van Wallendael (Ghent University - iMinds), Bert Vankeirsbilck (Ghent University - iMinds)

Goals

The VQEG Tools and Subjective Labs Setup group tries to bundle tools which can aid in the development of subjective quality metrics. These tools can be found on the following website: http://vqegstl.ugent.be/

Achievements 2014

This year, a new tool has been added, namely the **VQEGPlayer**. VQEGPlayer is a software for performing subjective video quality experiments for Windows 7 in 64bit. It is designed to present a video on a computer screen synchronized with the refresh signal of the display to allow for perfectly repeatable output if the PC system allows for sufficient performance.

Additionally, **MATLAB Code for Popular Subject Screening Algorithms** have been released on the website. The following popular subject screening techniques have been made available:

- ITU-R Rec. BT.500 Annex 2 Clause 2.3.1 recommends a technique for screening Double Stimulus Impairment Scale (DSIS) and Double Stimulus Continuous Quality Scale (DSCQS) data.

- ITU-R BT.1788, also known as SAMVIQ, demands that subjects have a stable and coherent method to vote degradations of quality.

- The Video Quality Experts Group (VQEG) HDTV Phase I Test Plan includes a method for screening subjects in Annex I (page 37).

- The Video Quality Experts Group (VQEG) Multimedia Phase I Test Plan includes a method for screening subjects in Annex VI (page 57).

Finally, in the list of recently contributed tools, the **Web-Enabled Subjective Test (WEST)** software package can be found. This package offers a solution to the problem of gathering subjective testing data from multiple locations and multiple portable or computing devices.

For completeness, the following tools which got contributed previous years are still provided on the website:

- Definitely Lossless: a solution to reliably transmit video data from one location to another

- H264AnnexBExtractor: a C++ application extracting the H.264 byte stream from network traces (.pcap files) containing H.264 streaming.

- PcapLossGenerator

- Sirannon: , formerly known as xStreamer, aims at being a modular multimedia streamer and receiver. The modularity is inspired by the Click Modular Router project an Direct Show filters.

- Telchemy PCAP/MPEG2-TS Scrambling Application: This application scrambles and unscrambles live streams and offline PCAP files.

- Telchemy PCAP Loss Insertion Tool: This software introduces losses to a pcap capture file using a 2-state or 4-state Markov model.

- IPTV-interface: Open source interface software for video quality monitoring. The monitoring of the service quality is crucial to the successful operation of commercial internet services to end customers.

- 3D Video Player: a full-featured 3D movie player.

- Modified JM H.264/AVC codec: Version 16.1 of the JM Reference software has been adjusted in order to enable the generation of an XML-based trace file.

Plans for 2015

For 2015, we are in the process of making available the **ITU-T P.1201 Audiovisual Quality Estimation Tool** implemented in Python. It estimates the audiovisual, video, audio coding quality for IP-based video streaming applications. It works on audiovisual bitstream (PCAP) in combination with the modified JM H.264/AVC codec. It also works with encrypted data, but the I/P/B frame sizes have in that case to be estimated, and the performance of the model depends on the accurate estimation of the I/P/B frame sizes.

Independent Lab Group (ILG)

Co-chairs: Phil Corriveau (INTEL), Margaret Pinson (NTIA/ITS)

Goals

The goal is to ensure that all VQEG validation testing is unbiased and done to high quality standards.

Achievements 2014

The ILG assisted in finalizing the Hybrid project.

Plans for 2015

Beginning in 2014 and extending into 2015, the ILG is helping the 3DTV project organize mature research on 3DTV subjective tests and pass this information to the ITU for inclusion in draft new Recommendations.

Joint Effort Group (JEG)

Co-chairs: Kjell Brunnström (Acreo Swedish ICT AB), Patrick Le Callet (IRCCyN)

Goals

Promotes the idea of joint collaboration within VQEG.

Achievements 2014

The number of JEG-related projects has increased and now all the projects are currently JEG.

Plans for 2015

Proposal is to change VQEG group names to reflect whether or not the effort is currently collaborative, through a “JEG-” prefix.

VQEG Administration and Web Support

Co-chairs: Kjell Brunnström (Acreo Swedish ICT AB), Margaret Pinson (NTIA/ITS), Arthur Webster (NTIA/ITS)

Goals

To take care of issues that do not fall within a VQEG workgroups and answering general questions about VQEG.

Achievements 2014

Planning and organizing 2 face-to-face meetings, one in Boulder, USA, Jan. 21-24, 2014, hosted by NTIA/ITS and Rohde & Schwarz and one in Stockholm , July 7-11, 2014, hosted by Acreo Swedish ICT AB.

The home page and mailing lists have been updated and maintained.

Plans for 2015

Two face-to-face meeting are planned, one in Santa Clara, USA, Feb. 23-27, 2015, hosted by INTEL Corporation and one not yet decided, but most likely in the United Kingdom in September.

eLetter

Co-chairs: Margaret Pinson (NTIA/ITS), Arthur Webster (NTIA/ITS)

Goals

The VQEG eLetter provides timely updates on recent developments, hot research topics, and society news in the area of video quality.

Achievements 2014

The VQEG eLetter was established in 2014, and published two issues. The first issue focused on training sessions for video quality subjective tests. The second issue focused on objective model verification and validation.

Plans for 2015

The VQEG eLetter board hopes to publish two issues of the eLetter in 2015. The first issue will be focus on consumer cameras.