



An Interview with Video Quality Experts

December 2009

An Interview with Video Quality Experts

Authors: Dr. Christian Schmidmer (Opticom), Dr. Akira Takahashi (NTT), Dr. Margaret Pinson (NTIA ITS), Dr. Stefan Winkler (Symmetricom), Kjell Brunnström (Acreo AB), Mr. Mu Mu (Lancaster University), Dr. Andreas Mauthe (Lancaster University)

By Mu Mu (Lancaster University)

Christian Schmidmer studied electronic engineering at the University Erlangen Nürnberg. After receiving his degree he worked in the audio group of the Fraunhofer HS in Erlangen for five years. His main research topics were audio coding and perceptual measurement. He is now the technical director of the Opticom GmbH, designing and selling the world's first PEAQ measurement system. He is CTO of Opticom since 1997.



Akira Takahashi received a B.S. degree in mathematics from Hokkaido University in Japan in 1988, M.S. degree in electrical engineering from the California Institute of Technology in the U.S. in 1993, and Ph.D degree in engineering from the University of Tsukuba in Japan in 2007. He joined NTT Laboratories in 1988 and has been engaged in the quality assessment of audio and visual communications. Currently, he is the Manager of the Service Assessment Group in NTT Service Integration Laboratories. He has been contributing to ITU-T Study Group 12 (SG12) on QoS, QoE, and Performance since 1994. He is a Vice-Chairman of ITU-T SG12, a Vice-Chairman of Working Party 3 in SG12, and a Co-Rapporteur of Question 13/12 for 2009-2012 Study Period. He received the Telecommunication Technology Committee Award in Japan in 2004 and the ITU-AJ Award in Japan in 2005. He also received the Best Tutorial Paper Award from IEICE in Japan in 2006, the Telecommunications Advancement Foundation Awards in Japan in 2007 and 2008.

Margaret H. Pinson earned a B.S. and M.S. in Computer Science from the University of Colorado at Boulder, CO in 1988 and 1990, respectively. Since 1988 she has been working as a Computer Engineer at the Institute for Telecommunication Sciences (ITS), an office of the National Telecommunications and Information Administration (NTIA) in Boulder, CO.

Stefan Winkler holds an M.Sc. degree in Electrical Engineering from the University of Technology in Vienna, Austria, and a Ph.D. degree from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is currently Principal Technologist at Symmetricom. Prior to that, he was Chief Scientist of Genista, which he co-founded in 2001. He has also held assistant professor positions at the National University of Singapore (NUS) and the University of Lausanne, Switzerland.



Dr. Winkler has published more than 50 papers and is the author of the book "Digital Video Quality". He also serves as an Associate Editor for IEEE Transactions on Image Processing. He has been an active contributor to the Video Quality Experts Group (VQEG) since it was founded in 1997 and is currently co-chair of the QoE Metrics Activity Group of the Video Services Forum (VSF).



Kjell Brunnström is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 20 years, including work in Sweden, Japan and UK. He has written a number of articles in international peer-reviewed scientific journals and conference papers, as well as reviewed scientific articles for international journals. He has been awarded fellowships by the Royal Swedish Academy of Engineering Sciences as well as the Royal Swedish Academy of Sciences. He has supervised a number of diploma work students. In the area of video quality assessment he has been active in the Video Quality Experts Group (VQEG) for many years and being a Co-chair of the Multimedia Group until the completion of phase I. Currently, he is Co-chair of the Joint Effort Group as well as the Independent Lab group in VQEG. Recently his research interests have been in video quality measurements in IP-networks and display quality related to the TCO-requirement.

Mu Mu received his bachelor's degree of Automation Engineering at Nanjing University of Aeronautics and Astronautics, China in 2002. His Masters degree of Science was awarded by the faculty of Electrical Engineering and Information Technology, Darmstadt University of Technology, Germany in 2006. He has been employed as a network security engineer in DJT-software, China working on bank terminal audit system and as a research student in Technology Centre, Deutsche Telekom, Germany working on the topic of Seamless mobility and QoS of next generation networks under the ScaleNet project funded by the German Ministry for Education and Research. Mu is in his final stages of his Ph.D at the Computing Department of Lancaster University, UK. His



research interests include multimedia networking, objective video quality assessment and management. His current research is supported by the European project - Network of Excellence CONTENT and Agilent Laboratories UK.



Andreas Mauthe is a Senior Lecturer at the Computing Department, Lancaster University. He has been working in the area of distributed and multimedia systems for more than 15 years. His particularly interests are in the area of content management systems and content networks, large scale distributed systems, peer-to-peer systems, and self-organisation aspects. Prior to joining Lancaster University, Andreas headed a research group at the Multimedia Communications Lab (KOM), at the Technical University of Darmstadt. After completing his PhD in Lancaster in 1997, Andreas worked for more than four years in different positions in industry. He was General Manager UK Operations, Chief Development Officer (CDO) and member of the division's management board of the Content Management Systems Division of Tecmath AG (now Blue Order), a German based software house and system integrator working in the area of content management (mainly for the broadcast industry). Andreas has been acting as organiser, chair, and programme committee member for various conferences. He is on the Editorial Board of the ACM Multimedia Systems Journal. Further, he has been participating in standardisation activities (e.g. ISO, SMPTE and VQEG) and served as expert advisor and evaluator for the European Commission.

Through previous studies in the field, SIGMM members have recognised that research in multimedia must consider aspects of human perception and user preference. Ultimately the improvement of the user level quality-of-experience is more desirable than the more easily measurable quality-of-service.

Today in the development of algorithms or systems researchers rely upon the use of existing video quality evaluation metrics to assess whether they can increase or maintain a certain quality-of-experience. Unfortunately, it is often unclear whether such metrics are actually applicable in a particular scenario or whether they have limitations with respect to numerous factors such as encoding format, environmental conditions, display device, to name just a few.

When Mr. Mu from Lancaster University proposed an article about video quality evaluation for the SIGMM Records, the editors took this as an opportunity to request expert advice that would provide a state-of-the-art discussion to the SIGMM Community. A catalogue of questions was prepared and sent to several renowned specialists in video quality assessment. We are grateful to Dr. Christian Schmidmer (Opticom), Dr. Akira Takahashi (NTT), Dr. Margaret Pinson (NTIA ITS), Dr. Stefan Winkler (Symmetricom) and Dr. Kjell Brunnström (Acreo AB) for their contribution to this article. We also thank Dr. Andreas Mauthe for his work in summarising the response of the experts.

Q1: Do you think that the current video quality models are robust enough for multimedia research in general?

Summary

The experts agree that most models are robust and perform well for the tasks they have been designed for. Specifically for the models standardised in ITU Recommendations such as J.247 it is acknowledged that they work well and without serious limitations. However, the performance of a model is heavily dependent on its application and the context in which it is used. For instance the use case scenarios as well as test scenarios have to be precisely specified and match the criteria the models have been standardised for. Thus, the successful use of the models in a research context very much depends upon their correct deployment in accordance with their given specification.

Christian Schmidmer

There is a multitude of models available in the market, but only very few of them have been thoroughly tested by independent labs on data that were not used for the training of the models. The best of these models are today standardised in J.247, which includes four different models. Two of these models have proven to be very reliable over a very large set of video data of different resolutions and with a very broad range of codecs and transmission errors (the subjective scores given to roughly 8200 file pairs were compared to the results obtained by the models. For details see www.vqeg.org [<http://www.vqeg.org>], Multimedia Phase I Report). Of course, all models have outliers, as do human listeners. If robustness is the key criterion, then PEVQ is the best tested model since it has the least outliers (in terms of minimum correlation under worst case conditions). As long as a user sticks to such models and views the measurement results with some healthy scepticism - as it should be good engineering practice - then at least PEVQ is good enough for multimedia research in general. It is however important to perform not only one measurement. We clearly recommend using a larger set of files and performing the measurement on those. When the measured difference between two versions of e.g. a codec becomes very small, then the results should be validated visually. Also, when decisions involving large investments have to be taken, then the final judgment must be based on subjective tests. However, massive measurements before that step can rule out many options, broaden the knowledge used for the decision taking significantly and reduce the amount and thus cost of required subjective testing dramatically. Important is also the usage of objective methods for quality monitoring. In this case all models standardised in J.247 can be used without serious limitation. All this is valid for those models that were tested by VQEG and which are standardised as ITU-T J.247. Their pros and cons are well known today and reported. The use of models not standardised by the ITU is clearly discouraged since the performance of these models has not been validated as thoroughly as for the standardised models.

Akira Takahashi

As for the standardized models such as ITU-T Rec. J.144 and J.247, they work sufficiently well for estimating subjective quality. However, one needs to be very careful about the quality factors and conditions on which these models were validated. For example, J.247 models were validated against QCIF-VGA videos, and should not be applied to, for instance, HD videos. The scopes of these Recommendations are clearly defined in the documents. On the contrary, I have serious concerns in the fact that many people use non-standardized models, which have not been validated by third parties, in their research. Or, applying the standardized models wrongly.

Margaret Pinson

Yes. The best proof that objective video quality models are robust enough for general use is an ANSI or ITU standard. The Video Quality Experts Group (VQEG) was established to perform independent validation of objective video quality models, which would serve as a basis for standardization. VQEG completed a validation test on video quality objective models suitable for multimedia applications in 2008. The VQEG Multimedia Final Report describes the performance of 25 models. These models were extensively tested using a variety of bit rates, coders, and packet loss levels. As a result of this validation, two ITU-T standards were approved: ITU-T J.246 "Perceptual audiovisual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference", and ITU-T J.247 "Objective perceptual multimedia video quality measurement in the presence of a full reference". The nineteen models included in these two standards are all well suited for many areas of multimedia research. The problem arises with research topics such as tweaking coding parameters slightly, because none of the objective models have the required accuracy. The VQEG Multimedia Final Report contains detailed analysis of model performance that will be of interest to multimedia researchers (without the need to understand the model details).

Stefan Winkler

To a large extent, the robustness of quality models depends on the application. It's important to know what artifacts, codecs, bitrates, quality range, etc. a given model has been tested on, and to apply it in that same context. Using a model designed for mobile video transmission to optimize a noise reduction algorithm without verification testing is risky. The impact of temporal conditions such as packet loss, packet-delay variation, clock synchronization, and so on, is not always captured well by video quality models. In general, it is important to be aware of the model's performance, accuracy, and limitations.

Kjell Brunnström

No not in general. If you limit the scope to the top scoring models in VQEG multimedia test and stick to the scope of those models in may be used (VQEG, "Final Report From the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase I", VQEG Final Report of MM Phase I Validation Test, Video Quality Experts Group (VQEG), (2008)). In VQEG Multimedia phase I multimedia was defined as small format video (QCIF, CIF and VGA) without audio, with a bitrate less than 4 Mbps. The models were tested on a variety of codecs and some transmission errors, using a large number of viewers in labs all over the world. There were three different model types that were evaluated: full reference (FR), reduced reference (RR) and no reference (NR). If the top scoring FR and RR models are used within the tested context then they are robust. The NR models are still not mature enough. The problem is that it is very likely that the models will be used outside this scope and then the result could not be trusted as predictor of perceived video quality.

Mu Mu

Video quality models were commonly designed with aspects of various quality degradations in the distributed system and aimed at dedicated assessment scenarios. Several models have been verified and recommended by independent groups such as VQEG using standardised validation tests. However, these validation tests only verify the performance of the models with a dedicated test plan (e.g. SDTV) which the models are designed on. How well the models will perform on other test plans is unknown and must not be taken for granted. It is also observed that most of the current objective models were intended for off-line quality evaluation. Complex image processing is required on the actual video frames and the original content must be available for the full reference models. With the advent of commercial multimedia services over packet networks, a real time video assessment service is becoming a fundamental requirement. However the light-weighted non-intrusive models which support in-service evaluation are still missing.

Q2: What should multimedia researchers that only use video quality models (without understanding their details and implications) be aware of?

Summary

The models have to be carefully selected and be appropriate for the test case. The users have to be aware of the artifacts, codec bitrates, etc. a given model supports. Further, the users should make an informed choice regarding the robustness of the model. As a guideline, standardised models can be used since they have been already validated.

Christian Schmidmer

- Use the best and most robust model you can get!
- Stay with international standards, others have already validated these for you!
- Don't trust results blindly. When in doubt, always perform a short visual validation!
- Don't trust a single measurement point!

- Know the limitations of the model!

Akira Takahashi

As written above, they should be aware of the scopes of Recommendations and conditions on which they were validated.

Margaret Pinson

The first and most important issue is to find independent validation of the model's performance. This is the key issue that concerns many consumers - can we trust this model? The models that have been standardized by ANSI, ITU-T, or ITU-R have undergone independent validation by the Video Quality Experts Group (VQEG, www.vqeg.org [<http://www.vqeg.org>]). This is ideal - independent proof that the model performance is sufficiently high as to be useful to consumers, conducted under carefully controlled circumstances with oversight. These standards will typically include details of the model's accuracy as measured by the validation testing, and the appropriate scope (e.g., appropriate applications). Where a standard is not available for a model, there may be conference papers published by a university that will yield insights into the model's performance.

The second issue is to understand that the best objective video quality model is not as accurate as subjective video quality testing. The models available today are, as a very rough analogy, as precise as one to two viewers within a very carefully conducted subjective experiment (as compared to a panel of 20 to 30 viewers in a typical subjective experiment). That is not the same thing as "one person watching the video", of course, because subjective testing contains many safeguards that remove viewer bias from the data.

Thus, the user should inspect paperwork describing the model, to get a feeling for the model's accuracy. One way is to visually examine scatter plots of the model's performance on testing data or validation data (i.e., video sequences that the model was not trained upon). One way to measure the confidence interval of a model's objective values has been standardized in ATIS T1.TR.72-2003 "Methodological Framework for Specifying Accuracy and Cross-Calibration of Video Quality Metrics," available at <https://www.atis.org/docstore/product.aspx?id=10518> [<https://www.atis.org/docstore/product.aspx?id=10518>], and ITU-T Recommendation J.149 (03/04), "Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)," available at <http://www.itu.int/rec/T-REC-J.149/en> [<http://www.itu.int/rec/T-REC-J.149/en>]. This method is demonstrated in the NTIA Technical Report "Techniques for Evaluating Objective Video Quality Models Using Overlapping Subjective Data Sets," available at www.its.bldrdoc.gov/pub/ntia-rpt/09-457/ [<http://www.its.bldrdoc.gov/pub/ntia-rpt/09-457/>]. The key lesson is that the user should acknowledge in their experiment design that the objective video quality model will be one source of error in their experiment.

The third issue is that all models appear to become more accurate when the results of several different video sequences are averaged together (i.e., the models appear to track average video system quality better than the quality of an individual scene). The scenes should be carefully selected to span a wide range of spatial, temporal, and other video characteristics (e.g., brightness, contrast, color). The NTIA document mentioned above demonstrates this phenomenon. Generally, better objective to subjective correlation results are obtained as the number of scenes are increased, but there are diminishing returns after about 15 video sequences.

The fourth issue is that all models will perform best for the conditions they were trained to handle. The model's documentation should indicate the conditions for which the model is intended. Some users will always need to deviate from this intended application; just be aware that the model accuracy is probably decreasing. The further you go from their intended use, the less accurate the model may become. For example, a model that is only trained on QCIF video only will probably be less accurate for CIF resolution video, and may yield misleading results for HDTV.

Stefan Winkler

It's important to know what artifacts, codecs, bitrates, quality range, etc. a given model has been tested on, and to apply it in that same context. Using a model designed for mobile video transmission to optimize a noise reduction algorithm without verification testing is risky. The impact of temporal conditions such as packet loss, packet-delay variation, clock synchronization, and so on, is not always captured well by video quality models. In general, it is important to be aware of the model's performance, accuracy, and limitations.

Kjell Brunnström

Use models that have been independently tested e.g VQEG and only use them within the tested scope.

Mu Mu

To justify any conclusion on perceived quality of video content, one either needs to conduct subjective user tests and provide statistical verification or employ an appropriate and good objective model. In the latter case, the limitation of the model must be investigated. Although some objective models are proven to be superior in some test conditions, they may produce inconsistent or even contradictory results under different conditions.

Q3: Do you think that existing objective video quality models are robust enough to be used by non-specialists?

(i.e. are objective video quality models sufficient for non-specialists who use them in their own field of expertise, or should they person subjective studies?)

Summary

Whereas the models themselves are perceived as robust there are issues ranging from the configuration of the systems to the right interpretation of the results. For example, in Full-reference models it is crucial to align the reference and degraded videos in both the time and spatial domain. Inadequate configurations of models can lead to false results. Objective models can be used as a reference of how certain factors affect the QoE of a video system. However, users should not consider the quality model as a universal tool which precisely quantifies the quality level regardless of context.

Christian Schmidmer

The best available models certainly are. Their results are reliable and the models are simple to apply. A general understanding of the measurement principle should however be available. The interpretation of the measurement results however, is a different thing. Here, detailed knowledge of the system under test is clearly required - but that's the case independent of the measurement method. In general, the usability of the model by naive users (naive as far as the measurement algorithm is concerned) is a matter of the testsystem design as well as the measurement algorithm being used. Systems that require many settings are of course a disadvantage here. Luckily most standardised algorithms don't require many settings. PEVQ for example simply takes two files of whatever framerate and image size and compares them. Of course it makes very little sense to compare a 1080p30 video with the same video in QCIF resolution at 4fps, but as long as the dimensions are about right, PEVQ will do the job. There is no need for a lot of preprocessing and no decisions have to be made about viewing conditions etc. Systems offering too many parameters may easily be misadjusted to predict whatever score the user wants to see....

Akira Takahashi

It depends on how they are implemented. For example, in Full-reference models, it is crucial to align the reference and degraded videos in both the time and spatial domains. If a product cannot cope with this well, the result is not reliable at all. This is just an example.

Margaret Pinson

Yes, there are many objective video quality models available today that can be used by non-specialists. All of the models identified in ANSI, ITU-T, and ITU-R standards fall into this category. At this date, this includes full reference (FR) models and reduced reference (RR) models. VQEG testing indicated that two no-reference (NR) models for some QCIF resolution video multimedia applications can be useful, but the owners of those models have not pursued a standard at this time. The non-specialist should apply the model in accordance with the approved usages that are specified in the respective standards and recommendations.

Stefan Winkler

Among our customers and model users, I've encountered many misunderstandings about video quality models. Among the most common ones are, for example, reporting MOS values with 5 decimal levels, or choosing a "better" video or system based on tiny MOS differences. Also, the big influence that the specific video content can have on model output is often neglected - for example, people often compare low-complexity with high-complexity video and are surprised about quality differences. Similarly, the impact of screen size, video resolution or frame rate is not well captured by many models, yet people like to use them for direct comparisons of mobile video with HD video quality.

Kjell Brunnström

Yes, some models are robust enough, but must be used within its scope, as explained above. However, this requires an easy to use mature implementation of the model. The software should also assist in interpreting the results, which could be hard for a non-specialist. It should be noted that performing a subjective study is not an easy task to perform, if it should be done right. It would also be very time consuming and comparably expensive.

Mu Mu

Video quality models have been used by non-specialists to verify their network or application design with the aspect of "QoE". It is worth noticing that every quality model has its limits and restrictions. Valid test results can only be achieved if the test scenario and procedure meets exactly the specification of the model. Choosing more than one model and comparing their results is recommended. A subjective test would always be helpful to validate any conclusion.

Q4: How do you see the role of subjective quality assessment for the development of objective video quality evaluation?

(i.e.: is it worthwhile showing Foreman to 5000 people and what is your opinion on using non-standard clips for quality assessment?)

Summary

Subjective quality assessment remains the benchmark and is the most essential, fundamental, and reliable way to quantify video quality. It has been found that even relatively small groups (around 30) give reliable and repeatable results. To test videos on a large number of users is considered inefficient and a waste of resources. Especially since the amount of work associated with conducting a subjective experiment is considerable.

Christian Schmidmer

Standard clips are the worst one can use since most systems will be trained on them. It makes no sense to use them for development of transmission systems or even worse, video quality measurement algorithms. Doing so will lead to overtrained systems that do not generalise well. As far as the algorithms are concerned, subjective tests are the most important basis for this development and such subjective test results are invaluable for the developer of measurement algorithms. The ultimate goal is of course to replace the need for subjective testing entirely, but this development is still far, far away. Also, as new transmission systems are developed, new distortions will be introduced into video signals and perceptual video quality models must be re-evaluated for their fitness to predict such distortions correctly. This can only be done by comparison to subjective test results. In short, subjective testing will still have a very long live...

Akira Takahashi

Subjective quality assessment is the most essential, fundamental, and reliable way to quantify video quality. However, it is quite difficult to plan and conduct a reliable subjective experiment. So, for those who are not well-skilled in this field, objective quality assessment may be appropriate if the aim of an experiment fits the scope of Recommendation. As for clips, they should use standardized ones if they would like to compare the result with other studies. The choice of clips is very critical in subjective (or objective) quality assessment. It is even easy to "manipulate" the evaluation results by choosing certain clips so that one's codec looks very good!

Margaret Pinson

Subjective quality assessment remains the "gold standard," as it is significantly more accurate than any objective model. Recent VQEG lab-to-lab results show that subjective testing is highly repeatable (see the NTIA report mentioned above). The accuracy of subjective scores increases as the square root of the number of viewers, so after a point there are diminishing returns. Thus, you need four times the number of viewers to tighten the confidence interval of the mean opinion score by a factor of two.

The ITU standards require a minimum of 15 viewers. In our experience there is little point in going beyond 20 to 30 viewers per clip. Showing one video sequence to 5,000 people would be a tragic waste of time and resources that could be better spent analyzing a much larger variety of source video sequences and impairments. The variety of scenes and impairments included in a subjective test are inevitably the controlling factors that define that experiment's accuracy. The scenes chosen are by necessity a small sample of all available content. If that sampling fails to span a sufficiently wide range of content, then the subjective results will be skewed, biased toward this particular set of scenes instead (as intended) to be representative of the wider range of all content. Someone with resources sufficient to show Foreman to 5,000 people would be much better served to show 25 different scenes run through 10 impairments each to 20 people. Another limiting factor of subjective tests is viewer pool and laboratory biases. This is one reason why the absolute quality rating of a video clip can vary considerably from laboratory to laboratory, but the relative quality ratings between the video clips are very stable.

The ITU and ANSI standard test sequences are valuable because these high-quality sequences are in the public domain. Some of these are available on the VQEG website (www.vqeg.org [<http://www.vqeg.org>]). There is a sense that these scenes were carefully selected, which is true. However, the standardization depended primarily upon availability of high-quality content that could be put into the public domain. Non-standard source sequences are equally valuable and appropriate, depending, of course, upon whether you can obtain a high-quality original recording and permission to use that content for the tasks at hand. Most researchers have extreme difficulty in obtaining high-quality source sequences. The Consumer Digital Video Library (CDVL) is being established to address this critical industry need. When CDVL is online in late 2009, this web site (www.cdvl.org [<http://www.cdvl.org>]) will let researchers and developers have royalty-free access to high-quality source sequences.

Stefan Winkler

Subjective quality assessment remains an essential benchmark for objective quality metrics. The amount of work associated with conducting a subjective experiment often intimidates people, and there is no doubt that this takes much preparation, but it is better to do a small informal subjective test than none at all!

It is also important to choose content that is both representative for the application at hand and covers a wide range of complexities. Unfortunately, the "standard" test clips available (Foreman) do not always fulfil these criteria. Other than the (important) option of sharing the database perhaps, there is no reason to limit oneself to those. The availability of good-quality source content remains an issue, particularly for HD video. With the wider availability of HD cameras, the situation is bound to improve, however.

Kjell Brunnström

Subjective tests are essential for the development of models. There should be a large variety of contents (not only Foreman to many subjects) crossed with error conditions. The problem is usually that conclusions of performance are drawn on too few subjective tests. Furthermore, it is quite common to publish results where the subjective data has been used both for training and evaluation, leading to misleading results.

Mu Mu

Subjective quality assessment is irreplaceable to study human users' perception on video content. Several international standards give recommendation on conducting valid subjective experiments. This includes methodologies on collecting user score, setting up

viewing conditions, selecting test materials as well as communicating with subjects (human participants). The number of subjects should be statistically sufficient to secure the significance of any outcome. However, the subjective test is usually conducted with limited resources in terms of time and budget. In most cases, the number of participants has to be limited to increase the efficiency of a test.

Q5: Which research steps in subjective and objective video quality evaluation do you expect in the next 5 years?

Summary

The development will reflect the upcoming coding deployments and research such as HD and potentially 3D. Further, it is expected to reflect different application domains such as public safety applications and medical applications. Another development will be the deployment and use in in-service scenarios, i.e. the models will be integrated into a service infrastructure. Finally, it is expected that new metrics will be developed to better reflect the user experience and new models will be researched.

Christian Schmidmer

- Models for HDTV.
- Audiovisual Models (both subjective methods as well as objective models).
- Hybrid models (Hybrid models combine the analysis of bitstreams with the analysis of the payload. This may for the first time lead to reliable no-reference measurement algorithms.)

Akira Takahashi

There are several directions, I believe. The first one is to extend the scope of current technologies, for example, from SD to HD. Another direction is to develop methods that can be used in in-service scenarios, in which available information and computational capacity are limited (i.e., cannot use pixel data in objective measurement). We also expect the studies on subjective and objective assessment of 3D videos. Cause-analysis can also be a topic.

Margaret Pinson

VQEG is currently conducting validation tests on HDTV models. The HDTV Final Report and resulting ITU standards are expected within the next year. The next VQEG validation experiment will likely examine hybrid models - that is, models that examine both the encoded bit stream and the decoded video as seen by the viewer. The validation results for Hybrid models and resulting ITU standards should be completed within the next few years.

The Public Safety Communications Research Program (PSCR, www.pscr.gov) program is investigating minimum performance criterion needed for various public safety applications (e.g., fire fighters, police, and emergency telemedicine). Their goal differs from traditional video quality subjective testing. This led to the development of task-based subjective testing, which investigates whether the quality of video is sufficient to perform a particular task (e.g., can you read a license plate at this level of compression?). There is a new standard that defines how to do these subjective tests: ITU-T P.912 "Subjective video quality assessment methods for recognition tasks." The PSCR initiative Video Quality in Public Safety (VQiPS) is developing specifications that work to improve the way in which video technologies serve the public safety community. VQiPS expects to release documents specifying objective ways to specify video quality requirements for these applications in the next few years.

Stefan Winkler

We all know that the still-ubiquitous PSNR is not a good quality metric in most cases. I hope in the near future we can finally replace it with better models, at least for certain applications. That will only happen if these new models are well-understood, easy to use, and accurate.

Certain "light-weight" models will find applications in end-user equipment, such as set-top boxes or even mobile devices, where a small footprint is more important than high accuracy. In practical situations such as distribution of multimedia content over packet networks, no-reference models will be required for real-time monitoring.

The most demanding applications are in the area of active control and optimization of video processing systems such as encoders, multiplexers, filters, etc. This I believe will require metrics that measure not only overall quality, but also give more fine-grained assessment of the video through metrics for specific relevant artifacts.

Kjell Brunström

How to systematically subjectively test the overall quality experience of multimedia taking into account the different modalities involved. This would lead to a base of developing objective quality models for that. There will also be a development of testing for 3D quality of experience.

Mu Mu

Most of the existing quality models were designed to evaluate compression design of video codec or error resilience of delivery mechanisms. The real time assessment was not taken as a requirement of objective models since evaluations were usually conducted before or after the services. With more high quality premium video services being delivered over packet networks, in-service quality monitoring has become an essential prerequisite

for quality assurance. A light-weight model which supports instant assessment over large quantities of video streams will be a highlight of future research in the field.

Q6: How can members of SIG Multimedia contribute to that research?

Summary

The biggest contribution could be achieved by publishing high-quality test material and sharing subjective test-results that can be used to validate objective models. Further, joint work on specific models and contribution to standards would be useful.

Christian Schmidmer

By publishing subjective tests, by participating in VQEG, by presenting educational information in journals etc.

Akira Takahashi

In some cases, objective models need to know the details in coding algorithms. So, contribution on the analysis of the effects of each coding parameter on the video quality is very helpful.

Margaret Pinson

One key bottleneck to all video quality objective model research and validation is the lack of high quality source video sequences. SIG Multimedia can contribute to this research by making their high quality video sequences available to other researchers and developers through the CDVL web site, www.cdvl.org. This is an inexpensive way to promote research and standardization.

Another way SIG Multimedia members can contribute is to support the VQEG Independent Lab Group (ILG). The task of the VQEG ILG is to provide independent oversight and validation of objective video quality models. Tasks include providing secret content (previously unseen by the model developers), creating impaired video sequences to serve as test vectors, and running viewers through subjective tests. The support of organizations to help independently validate models will be particularly important for the Hybrid validation effort. Please contact the VQEG Chairs, VQEG ILG Chairs, or Margaret Pinson for more information (see www.vqeg.org [<http://www.vqeg.org>]).

There are many opportunities for research and thesis topics delving into quantifying the quality of video needed for public safety applications. The public safety video sequences being made available on CDVL will facilitate these avenues of research possible. For more information on the Public Safety Communications Research program, go to www.pscr.gov [<http://www.pscr.gov>]

Stefan Winkler

One of the best ways to advance the state of the art is sharing video quality databases annotated with subjective ratings, and sharing objective models. Open databases allow researchers to benchmark and compare their models easily, and open models greatly increase their chances of being used and tested in various applications. This open approach has helped the popularity of SSIM and the LIVE image quality database, for example.

Also, if you can, contribute to standards, even though I am aware that the membership fees are sometimes prohibitive to academic institutions or smaller companies. There is some very interesting work going on currently in various groups, such as proposals for a different model evaluation process (ATIS IIF), efforts to develop quality models by collaboration (VQEG or ITU-T SG12), etc. VQEG in particular is a great informal group to participate in, as it is open to everybody.

Kjell Brunnström

I am not so familiar with the work of SIG Multimedia. In general, there is a need to take the research further for developing models for multimedia. For that the increased understanding of the basis human judgement of quality is needed.

Mu Mu

The quality assessment is a topic which covers several research domains such as human psychology, image processing, video codec, networking and statistics. A successful model design requires expertise of all these different domains. For instance, studies on the network impairment patterns in practice networks can help researchers to design adequate models that perform well in realistic conditions. The contribution can come from research publications or exchanging experience in international study groups. In VQEG, a joint effort group has recently been initialised. The goal of this group is to work on a defined quality model using the experience and knowledge of group members.

[Previous Section](#)
[\[awardopps.html\]](#)

[Table of Contents](#)
[\[highlights.html#toc\]](#)

[Next Section](#)
[\[phd_students.html\]](#)
