VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice Habilitační a inaugurační spisy, sv. 537

ISSN 1213-418X

Martin Slanina

QUALITY OF EXPERIENCE IN THE MULTIMEDIA SYSTEMS OF TOMORROW

BRNO UNIVERSITY OF TECHNOLOGY
Faculty of Electrical Engineering and Communication
Department of Radio Electronics
Field of Research
Electronics and Communication

Ing. Martin Slanina, Ph.D.

QUALITY OF EXPERIENCE IN THE MULTIMEDIA SYSTEMS OF TOMORROW

KVALITA ZÁŽITKU
V MULTIMEDIÁLNÍCH SYSTÉMECH ZÍTŘKA

SHORT VERSION OF HABILITATION THESIS

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ NAKLADATELSTVÍ VUTIUM

BRNO 2016

Keywords
Quality of Experience, multimedia, video coding, stereoscopic display, mobile video, HTTP Adaptive Streaming, temporal pooling, Peak Signal to Noise Ratio, Artificial Neural Network, Scalable Video Coding.

Klíčová slova
Kvalita zážitku, multimédia, kódování videa, stereoskopické zobrazení, mobilní video, adaptivní streamování přes HTTP, špičkový poměr signál-šum, umělá neuronová síť, škálovatelné kódování videa.

MÍSTO ULOŽENÍ HABILITAČNÍ PRÁCE

Práce je uložena na Vědeckém oddělení děkanátu Fakulty elektrotechniky a komunikačních technologií Vysokého učení technického v Brně, Technická 10, 616 00 Brno

ISBN 978-80-214-5347-0
ISSN 1213-418X

Contents

1. About the Author

	Ing. Martin Slanina, Ph.D.
born	April 15, 1982
e-mail	slaninam@feec.vutbr.cz

Current position

since 2008	*Assistant professor* at FEEC BUT with specialization in digital television technology, video and multimedia technology, video quality, video signal processing and mobile communication
since 2015	*Director* of SIX Research Centre (Sensor, Information and Communication Systems)

Qualification

2009	Ph.D., Electronics and Communication (BUT)
2005	Ing., Electronics and Communication (BUT)

Research

Video and multimedia quality assessment, video coding, multimedia services in mobile communication systems
Papers indexed by Thomson Reuters:
http://www.researcherid.com/rid/E-9452-2014
Papers indexed by Google Scholar:
http://scholar.google.cz/citations?user=QLGPUnEAAAAJ&hl=en

Internships

2012	Telecommunications Research Center Vienna FTW (Austria) – 3 months
2012	Blekinge Tekniska Hogskolan (Sweden) – 2 weeks
2011	VTT Research Centre of Finland, Oulu (Finland) – 3 weeks
2006	Linköping University (Sweden) – 2 months

Personal skills

Languages	Czech – mother tongue English – fluent, FCE certificate German – basic knowledge

2. Introduction

"Luck is what happens when preparation meets opportunity..." The Roman philosopher Seneca very wisely put the two basic building blocks of success together in mid-1st century AD already. For sure it is daring to speak about the multimedia systems of tomorrow before they are deployed or even standardized. Some of the systems may not live to get beyond a concept or proven concept phase. But let's get prepared for the case they are successful at last!

The *multimedia systems of tomorrow* in the context of this thesis should be understood as multimedia services which are in early deployment phase or likely to be deployed in near future, no matter if they will be here to stay. This thesis includes scientific work considering both the successful an the recessive approaches.

The limits of knowledge can only be pushed if the state of the art knowledge is understood. Once understood, the knowledge can efficiently be provided to interested learners. The educational aspect is thus an inevitable part of the academic professional work and is also considered in this thesis.

2.1 Current Trends in Multimedia Services

The way people are consuming multimedia has changed. Twenty years ago, television broadcasting was the dominant way to reach a movie, live football match broadcast or video news at home. Television broadcasting is a linear service whose audience has quite limited decision power – zap to a different channel or record the broadcast locally for later viewing. No interactivity, no rich experience, no activity was required. Since then, the main improvement we could experience lies in the vastly improved flexibility [5]:

- Linear television broadcasting has been extended by connecting with online streaming for on demand content availability, creating the Hybrid Broadcast Broadband TV standard (HbbTV) [4].
- Online streaming together with the deployment of mobile devices (smartphones, laptops, portable computers) and mobile connectivity have opened a whole new market in which "media anywhere, anytime" is the key driver.
- Stereoscopic (also Three-Dimensional, 3D) television has become an optional improvement of viewing experience both in the cinema and at home. Although the rapid popularity growth of stereoscopic displays is over, the technology will likely be here to stay and improve with new types of displays being introduced [1]
- The Ultra High Definition Television (UHDTV) is very likely to become a key technology in improving the home viewing experience [2] in the following years.

Obviously, all these trends bring new types of services and new challenges in assessing their performance. Assessment in terms of Quality of Service (QoS) is gradually shifting towards Quality of Experience (QoE) [3] since QoE involves not only different technical parameters and performance indicators, but also the impact of the consumer himself on the experience he receives. Quality of Experience in the emerging multimedia systems is the core topic of this thesis.

2.2 Background and Motivation

When I applied for the topic "Methods and tools for video quality assessment" in the Ph.D. programme of Brno University of Technology under the supervision of prof. Václav Říčný in 2005, I was hardly aware of the broadness of the problem, the different quality assessment / measurement approaches available and the number of issues waiting to be solved. My two-month internship at Linköping University in Sweden with prof. Robert Forchheimer and very interesting discussions with Johan Görsjö (Agama Technologies) helped me find my way and focus on no-reference video quality measurement based on bit stream analysis. The continuing work on the design of a no-reference objective video quality metric led to a successful defense of Ph.D. in 2009. Since the topic turned out to be interesting in the scientific community in the following years as well, several papers were published after the thesis defense.

The Ph.D. work focused on objective video quality measurement, i.e., automated quality analysis. To be able to benchmark any objective video quality measurement algorithm, a ground truth is needed as reference. Such ground truth is typically collected through a subjective scoring session, where a group of observers provides quality scores in a carefully designed and controlled experiment. Although the test procedures in such experiments are standardized for common scenarios, they do not always correspond to the use case of the service. For instance, test procedures for standard definition television systems may not be appropriate for high definition television systems or even stereoscopic television system under evaluation. Further, the controlled laboratory environment may not reflect the typical use case of a system (e.g. mobile multimedia in a cell phone or tablet). These issues and the desire to understand the laws of perception in stereoscopic television systems and mobile television led to focus on QoE in different viewing contexts.

Mainly due to the cost of content production and the related content availability, stereoscopic television never found its way into a majority of living rooms. On the other hand, the viewing habits of multimedia consumers are slowly changing. Online streaming is gaining increasing importance in both home environment (typically connected to fixed network connection, viewing on a television set) and on the go (typically connected to a mobile network, viewing on a mobile device). A typical problem in online streaming is that the available bit rate is changing over time, which leads to the risk of stalling (video freeze) in case the available bit rate is lower than media bit rate or network underutilization in the opposite case. Guided by the intention of designing a suitable objective quality metric, I started to examine the performance of scalable coding as a possible solution, spending three weeks in the VTT Reseach Centre of Finland in Oulu, making use of the chance to discuss with Janne Vehkaperä, Martin Varela and Toni Mäki. It turned out that scalable coding, at its state of development, offered rate-distortion efficient coding only in strongly constrained coding setups. Later on, I had a chance to spend several months at Forschungszentrum Telekommunikation Wien (FTW) with the group of Raimund Schatz, working on quality evaluation of multimedia carried through HTTP adaptive streaming.

The whole thesis is divided into five parts. The core part of this habilitation thesis is formed by peer-reviewed research papers published in international conferences and international journals spanning three topics of research: *Metrics for No-Reference Video Quality Measurement*, *Quality of Experience in Various Viewing Contexts* and *Scalable and Adaptive Video Streaming*. As networking and team work is extremely important in scientific work, I value most the work done in international research teams. Most of the papers included in this thesis are thus an outcome of cooperation in a group of international authors.

2.3 Bibliography

[1] Cohen, D. S.: 3D TV: Not Dead Yet, and Getting Better. 2014.
URL `http://variety.com/2014/digital/columns/3d-tv-not-dead-yet-and-getting-better-1201055163/`

[2] Joseph, J.; Cassagnol, D.: The Future is Clear – 4K Ultra High-Definition Continues to Build Positive Momentum. 2015.
URL `https://www.ce.org/News/News-Releases/Press-Releases/2014/The-Future-is-Clear-%E2%80%93-4K-Ultra-High-Definition-Con.aspx`

[3] Le Callet, P.; Moller, S.; Perkis, A.; et al.: Qualinet white paper on definitions of quality of experience. 2013.

[4] Merkel, K.: Hybrid broadcast broadband TV, the new way to a comprehensive TV experience. In *2011 14th ITG Conference on Electronic Media Technology (CEMT)*, March 2011, p. 1–4.

[5] Trautman, E.: 5 Online Video Trends To Look For In 2015. 2014.
URL `http://www.forbes.com/sites/groupthink/2014/12/08/5-online-video-trends-to-look-for-in-2015/`

3. Research Overview

In this Chapter, we will make an overview of the papers included in this habilitation thesis. In total, author's work published in 9 papers forms the core of this thesis, organized in three topics. All the included research work has been peer reviewed and published in international conferences (6 papers) or international journals (3 papers) in the years 2007 – 2015. The following Sections show the basic background behind each of the parts, reveal the ideas and summarize the most important results achieved in the included work. Last but not least, papers are linked together so that their continuity is made clear.

3.1 Metrics for No-Reference Video Quality Measurement

In the early days of the first television broadcasting, the main factor for judging the overall quality of the system at the receiver was whether or not the desired program was shown. Although often noisy and blurred, the video image quality was generally satisfactory as long as faces and motion were recognizable.

Since then, times have changed. People got used to viewing sharp colored images with smooth motion and fine details. Delivering such quality is nowadays a must, not an option. In order to make sure the customers do not complain about poor media quality, it is very important for the service provider[1] to know the characteristics of the media at the customer's end. This is, however, not an easy task in the systems using digital processing and transmission techiques as the final quality is highly content dependent.

Generally, the tools allowing automated measurement of signal / displayed quality are called *objective* techniques. On the other hand, *subjective* procedures collect quality rating from a group of observers, which is obviously an expensive and time consuming approach, not applicable in many scenarios such as in-service quality monitoring.

Objective media quality measurement techniques are usually classified with respect to the input data they require, reaching from *full reference* methods requiring access to a "perfect quality" reference for comparison to *no reference* methods relying on the received data available at the received end only. Obviously, different approaches are suitable in different scenarios. For instance, when a video encoder is being developed and optimized, it is desired that the uncompressed original signal is kept perceptually unchanged by the encoding. Full reference quality measurement models can then reveal the differences, i.e., show how much the compression impacts and degrades the sequence. In other scenarios, full reference quality measurement is not suitable and not even possible. For quality measurement at the receiving end of the system, the unmodified original content is naturally usually not available. Quality thus has to be judged based on nothing but the received (more or less degraded) signal. This is intuitively a more common and challenging scenario. The lack of full reference methods was the motivation for my research in this area.

[1]The provider may or may not be a single entity as a number of stakeholders may be involved in delivering the desired media to the user - e.g. content producer, network operator, etc.

The usual approach in designing an objective video (or audio) quality metric is trying to mimic the subjective scores – the quality ratings that a group of people would give to the assessed material. There are, however, several shortcomings that need to be considered:

- *The subjective scores are enumerated through statistics of ratings gathered from a limited number experiment participants, the generalization to a whole population is not straightforward and the subjective scores are thus ambiguous.* All people are individuals and the subjective ratings they provide are more or less different from person to person. To achieve the most accurate results, we would require a huge number of people providing scores, making statistics of a large set, which is obviously not feasible. Thus, we use statistics of a small group (usually up to 20 people) to estimate the properties of the whole population. Such generalization is generally done through parametric probability distribution model fitting, typically defining in which range the mean score of the whole population is with a given level of confidence.

- *When a parametric probability distribution model is used, the distribution of the selected data should be selected and tested properly*. This is, unfortunately, rarely done and for many databases of media data with subjective scores available, normal distribution is assumed without further testing. For categorical values, however, normal distribution is not ideal.

- *Gathering subjective scores for large sets of data is lengthy and expensive*. For a subjective experiment, it is needed to hire the people to take part in the scoring session, explain the session principle and organization, collect the votes and analyze the result.

- *It is difficult to analyze different parts in a sequence since subjective scores give dominantly a single number per sequence.*

In fact, there are several very popular and widely used full reference objective test methods, whose results are very well understood and give a good basic clue on the perceived quality of the material under test. Being able to estimate full reference objective score without access to the reference is obviously very useful and may well substitute the estimation of subjective scores. Our approach to the problem will be described in the following paragraphs.

The content, results and impact of these papers will be discussed in Subsections 3.1.1, 3.1.2 and 3.1.3, respectively. In all the papers, video bit stream analysis is used to estimate a selected full reference objective metric results.

3.1.1 A Novel Metric for H.264/AVC No-Reference Quality Assessment

Originally published as:
Slanina, M.; Ricny, V.; Forchheimer, R.: A novel metric for H.264/AVC no-reference quality assessment. In *14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, June 2007, p. 114–117, DOI: 10.1109/IWSSIP.2007.4381166.

Abstract:
The article desribes a metric for H.264/AVC compressed video quality evaluation, which is being developed by our group. The metric operates on the encoded bitstream only. It uses the data directly present in the bitstream (or just entropy coded). No pixel decoding should be necessary, which will result in low computational complexity, and cosequently the possibility to analyze several bitstreams simultaneously. In this article, we describe one part of the system which has already been implemented - the feature extractor. A simple analysis is done to demonstrate the variability of the selected parameters used for evaluation.

This is the first work defining our basic concept for quality estimation without reference. The concept is based on the following targets:

- Develop a no-reference quality estimation metric for the most widespread video codec, i.e., H.264/AVC [19];
- Allow for a per-frame analysis to detect locations of low quality parts in a sequence;
- Make the estimation simple and fast: use features available in the compressed bit stream as input data, so that no decoding is necessary and thus computational time is saved;
- As coding quality is mostly influenced by the encoder setup and by the content being processed, include features representing both parts (such as quantization parameter representing the encoder setup and prediction mode distribution resulting from encoder decisions based on content).

A detailed description of the concept can be found in *A Novel Metric for H.264/AVC No-Reference Quality Assessment* [51].

The impact of this conceptual work may be demonstrated on the citing papers. A similar approach to video quality estimation of compressed H.264/AVC bit streams was used by many authors in the following years. For instance, it was used in [54], where the authors use the compressed stream bit rate, quantization parameter and number of skipped macroblocks as features, i.e., inputs to the quality estimation algorithm. Classification using multi-linear regression was then employed to estimate subjective scores for sequences being analyzed. Keimel et al. describe a no-reference metric using bit stream analysis in [22, 21]. Their approach is aiming at high definition video content and subjective quality estimation. Compared to our metric, the authors use an extended set of bit stream features, including the number of bits per slice and average motion vector length per slice. Also a different classifier (Partial Least Squares Regression – PLSR) is employed. Shi et al. in [43] demonstrate a further improvement of model performance by a more thorough analysis of quantization parameter distribution. In [14], the authors extend the partial least squares regression method to support arbitrary length sequences by performing the PLSR analyses on a per-Group of Pictures (GOP) basis.

Hu and Jiang in [64, 65] present a metric for 4K-Ultra High Definition (UHD) content. Again, the metric considers H.264/AVC compressed video and uses bit stream features as input – in fact exactly the same features as those used in our model. The features are then weighted using a multi linear regression model in order to estimate coding Peak

Signal to Noise Ratio (PSNR). The set of features is further extended in [65], where the authors also estimate one additional objective metric values – the Structural Similarity Index (SSIM).

In contract to the methods mentioned so far, Wang et al. in [61] use PLSR to estimate the quality of an H.264/AVC-encoded material considering both the coding degradation and the transmission degradation. The set of coding features for quality estimation is thus extended with network-related features, including transmission bit rate, delay, jitter and packet loss. It is worth mentioning that the network features in [61] could be adjusted in the NS-2 simulator[2], but in real transmission, their continuous measurement on the fly is not trivial.

Finally, Shahid et al. in [42] include our concept in a survey paper on no reference objective image and video quality assessment approaches.

Overall, the concept has proven its usability for no reference estimation of objective full reference quality measures as well as subjective scores. Looking at the performance of the different metrics listed in this summary, the achievable Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Count (SROCC), as the most common estimation fidelity measures, are well above 0.8 or even 0.9, which expresses a very precise estimation. As will be further described in Sections 3.1.2 and 3.1.3, we were also able to achieve similar scores using the Artifical Neural Network (ANN) classifier for PSNR and SSIM, though earlier than the citing works.

3.1.2 Estimating H.264/AVC Video PSNR Without Reference Using the Artificial Neural Network Approach

Originally published as:
Slanina, M.; Říčný, V.: Estimating H.264/AVC video PSNR without reference using the artificial neural network approach. In *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, Porto, 2008, p. 244–250.

Abstract:
This paper presents a method capable of estimating peak signal-to-noise ratios (PSNR) of digital video sequences compressed using the H.264/AVC algorithm. The idea is in replacing a full reference metric - the PSNR (for whose evaluation we need the original as well as the processed video data) - with a no reference metric, operating on the encoded bit stream only. As we are working just with the encoded bit stream, we can spare a significant amount of computations needed to decode the video pixel values. In this paper, we describe the network inputs and network configurations, suitable to estimate PSNR in intra and inter predicted pictures. Finally, we make a simple evaluation of the proposed algorithm, having the correlation coefficient of the real and estimated PSNRs as the measure of optimality.

While the conceptual work presented in Sec. 3.1.1 described just a general approach, a complete no reference quality metric is described here. In the concept, several issues were left unanswered:

[2] `http://www.isi.edu/nsnam/ns/`

- Content type to be considered;
- The target metric to be estimated;
- The classification of features, i.e., how to map the features into quality measures;
- Learning and validation procedure.

These questions are aimed at in the paper *Estimating H.264/AVC Video PSNR Without Reference Using the Artificial Neural Network Approach* [48]. The following paragraphs discuss the respective items raised above.

Our concept relied on low resolution video sequences. For instructional purposes, these sequences offer high transparency and low complexity processing, but their importance in real life usage is decreasing gradually. Recall that even for mobile video services, high definition content is not an exception but rather a usual option. The metric thus uses only high definition video sequences, as a representative of high perceptual quality service demanded by consumers.

The requirement for the metric is to remove the necessity of having a reference video sequence for comparison. The motivation for selecting the Peak Signal to Noise Ratio was its extremely broad usage in the video signal processing community. It is often articulated that PSNR is not a good metric due to its poor correlation with subjective experiment results. This is undoubtedly true, but there are advantages to PSNR as well. Due to its wide deployment, it is well understood and its limits are well known. Furthermore, for expressing the quality of a compressed video sequence with respect to compression algorithm, PSNR is the metric of choice in a majority of scenarios.

It has been shown in Section 3.1.1 that different regression tools are usable for the mapping of features into quality measures. Our metric uses the Artificial Neural Network (ANN) machine learning tool as a simple to use yet powerful classifier. Since we have identified different sets of input parameters for the intra (reference) frames and inter (using motion estimation and motion compensation) frames in a sequence due to different sets of available prediction modes, we have built and tested two ANN setups.

Learning and validation is extremely important in machine learning in general. If poorly performed, the results may not be extensible for different content or for different processing parameters. In our PSNR estimator, we used two distinct content sets, one for learning and the other for validation. It is understood that this way the generalization ability of the trained classifier is well examined.

The attained linear correlation between the real PSNR values and the PSNR estimated without reference in the validation set of sequences was 0.97 for intra coded pictures and 0.93 for inter coded pictures. These results are generally similar to those achieved by PLSR regression by the authors referenced in Section 3.1.1.

3.1.3 Reference Free SSIM Estimation for Full HD Video Content

Originally published as:
Ries, M.; Slanina, M.; Garcia, D.: Reference free SSIM estimation for full HD video content. In *Radioelektronika (RADIOELEKTRONIKA), 2011 21st International Conference*, April 2011, p. 1–4, DOI: 10.1109/RADIOELEK.2011.5936447.

Abstract:
This paper proposes reference-free video quality estimation method for full high definition video services based on structural similarity index. Design of our estimator is based on an artificial neural network. To achieve this, the neural network was trained with a set of video statistical parameters extracted from the most representative video contents. Moreover, estimations with neural networks allow higher applicability and require lower processing power as known reference based methods. Finally, achieved correlation between calculated and estimated structural similarity index shows very good fit.

Finally, a different application of our no reference metric is presented in the paper *Reference Free SSIM Estimation for Full HD Video Content* [35].

This work includes several modifications of both the concept and the PSNR estimator metric presented in previous subsection. The modifications are as follows:

- Different content is used;
- Different set of features is fed into the ANN;
- The target full reference metric is replaced.

The content used for the paper was selected to reflect the most popular content viewed in online video services as closely as possible. Thus, instead of uncompressed test material from research video databases, transparent quality compressed content was used due to much broader possibilities of usable content types. The most popular genres in Europe, namely action movie and soccer, were then represented in the content used.

Further, we decided to select a wider set of features to make use of more information, available from the compressed bit stream. Apart from quantization and prediction related parameters already introduced in the conceptual work, we also employed the number of bits in a frame spent on coding prediction residuals, the number of bits spent on motion vectors description and the number of remaining bits required for representation of a frame. In other words, we used a detailed coding bit rate analysis including bit rate classification.

Finally, the PSNR metric was replaced with the Structural Similarity Index, which has shown higher correlation with subjective scores than PSNR. We have achieved linear correlation in the validation set above 0.91, which is slightly worse than the correlation achieved for PSNR, but still very competitive. The metric is identified as a well performing no reference quality measurement algorithm in a survey work by Wulf and Zölzer focused on quality measurement for high definition content [63].

A natural continuation of the work on no reference quality metrics is an extension to subjective quality scores as estimation targets, which is also the content of the following sections.

3.2 Quality of Experience in Various Viewing Contexts

In recent years we have witnessed a number of radical changes in the way multimedia are consumed in the variety of multimedia services offered. First such change was the transition from analog television broadcasting to DVB-based digital broadcasting, finished in the Czech Republic at the end of 2011. The transition to digital broadcasting allowed for broad distribution of digital high definition content, while at the same time improved Internet connectivity was an enabler for a number of services offering high definition online streaming. At the same time also smartphones and tables increased in popularity and became very important end devices for consuming multimedia. Further, stereoscopic displays started selling in Central Europe in 2010 and literally flooded the market with television receivers in the following years. Stereoscopy was introduced in cinemas in around the same time.

Obviously, a lot has happened around us in the field of multimedia during the last five years. Naturally, new services bring new challenges – in the light of QoE, the following issues attracted my attention most:

- *Applicability of different approaches to subjective quality testing methodology*. There is a number of different test procedures described in relevant ITU Recommendations (e.g. ITU-R BT.500 or ITU-T P.910). While the selection of the appropriate methodology is often motivated by the (un)availability of reference for comparison, there are also different time instants at which user opinions may be collected. Does it make a difference whether the observer is rating continuously during a sequence or just at the end?

- *Validity of standardized subjective test procedures when used with the new content types (high definition and stereoscopic)*. The standard test procedures specify viewing conditions and display parameters in detail, in order to assure that there is no impact of the actual setting on the test results. Such setup is, however, not realistic in many situations today. Moreover, the standard grading scales are limiting the answers to a single dimension, e.g., how good the quality is or how perceivable the degradations are. Especially for stereoscopic viewing, however, we should be interested in more than just the quality – also the feeling of presence in the scene is imporant, the naturalness of the stereoscopic sensation as well appearance of a viewing discomfort.

- *Appropriateness of subjective test procedures in the new context (viewing on different devices) and in different environment.* This item is very closely related to the previous one. First, standard test procedures do not take into account the new device types. Second, quite intuitively the perceived quality of the same content displayed on different devices with different screen sizes, resolutions, viewing distances and surrounding environment are very likely to differ. Can we tell how big is the difference?

The content, results and impact of these papers will be discussed in Subsections 3.2.1, 3.2.2 and 3.2.3, respectively.

3.2.1 Analysis of Temporal Effects in Quality Assessment of High Definition Video

Originally published as:
Slanina, M.; Kratochvil, T.; Polak, L.; et al.: Analysis of temporal effects in

quality assessment of high definition video. *Radioengineering*, vol. 21, no. 1, 2012.

Abstract:
The paper deals with the temporal properties of a scoring session when assessing the subjective quality of full HD video sequences using the continuous video quality tests. The performed experiment uses a modification of the standard test methodology described in ITU-R Rec. BT.500. It focuses on the reactive times and the time needed for the user ratings to stabilize at the beginning of a video sequence. In order to compare the subjective scores with objective quality measures, we also provide an analysis of PSNR and VQM for the considered sequences to find that correlation of the objective metric results with user scores, recored during playback and after playback, differs significantly.

The paper *Analysis of Temporal Effects in Quality Assessment of High Definition Video* [46] is focusing on the implications of the selection among available standard test procedures for a subjective quality experiment.

When planning for a subjective quality assessment experiment, the experimenter needs to select the appropriate methodology and be aware of the limitations thereof. Our motivation is closely related to the objective metric design presented in Section 3.1 – since the metrics are frame-based, also a time aligned subjective quality score for each frame would be ideal for precise verification of metric performance or for training the metric for subjective score estimation. Intuitively, the human observer does not react instantaneously and is unable to report a quality change at the same time instant as displayed on the screen. The questions to be asked are thus the following:

- What is the reaction delay in a subjective experiment with continuous score collection?

- What is the difference in results when a continuous score collection test procedure is used in contrast to a single grading after the end of sequence playback?

In order to answer these questions, a subjective test session has been organized. The reaction delay has been studied through rapid quality changes between two consequent sequences (simulated scene cut). We have found that the reaction times are in the order of seconds (up to approximately 5 seconds), which makes it difficult to find the exact time instant of a quality change in the sequence based on just the subjective scores. The results prove that score collection interval up to one second is sufficient in a continuous score collection experiment.

We have also found in this experiment that the continuous ratings collected over time tend to be strongly biased. Our recommendation and our direction in work following up is to rely on subjective methodologies which collect the quality scores from observers after the presentation of a sequence ended.

The findings published in the above mentioned paper inspired Montero et al. [30] in the selection of the appropriate subjective methodology for their experiment.

3.2.2 Testing QoE in Different 3D HDTV Technologies

Originally published as:
Slanina, M.; Kratochvíl, T.; Ricny, V.; et al.: Testing QoE in different 3D HDTV technologies. *Radioengineering*, vol. 21, no. 1, 2012.

Abstract:
The three dimensional (3D) display technology has started flooding the consumer television market. There are a number of different systems available with different marketing strategies and different advertised advantages. The main goal of the experiment described in this paper is to compare the systems in terms of achievable Quality of Experience (QoE) in different situations. The display systems considered are the liquid crystal display using polarized light and passive lightweight glasses for the separation of the left- and right-eye images, a plasma display with time multiplexed images and active shutter glasses and a projection system with time multiplexed images and active shutter glasses. As no standardized test methodology has been defined for testing of stereoscopic systems, we develop our own approach to testing different aspects of QoE on different systems without reference using semantic differential scales. We present an analysis of scores with respect to different phenomena under study and define which of the tested aspects can really express a difference in the performance of the considered display technologies.

While in the work mentioned in previous Subsection considers applicability of different standard subjective test procedures, the work published in paper *Testing QoE in different 3D HDTV technologies* [47] is targeting an extension of the standards subjective test procedures in order to get more detailed information about a stereoscopic video presentation.

The paper is aiming at several important issues in the quality assessment of stereoscopic video presentations:

- Definition of an appropriate test methodology, capturing more than just a quality rating from the observers;
- Comparison of different stereoscopic display technologies and different content characteristics on the overall Quality of Experience.

In stereoscopic television, the observers receive a richer sensation due do the depth illusion in the presented video. The depth illusion is achieved by presenting slightly different sequences to the left and to the right eye, corresponding to different natural views of eyes being several centimeters apart. Technically, the illusion is most commonly achieved by using a pair of glasses, capable of splitting up the sequences shown on a display in the spatial (typically passive systems with polarized glasses) or temporal domain (typically active shutter glasses with synchronized dimming). The sensation is, however, not exactly the same as when viewing a real scenery. The reasons include the disproportion of vergence and accommodation of the eyes, depth illusion strength influenced by the stereo base (camera distance) of the shooting system or temporal / spatial resolution decrease in different display technologies. For these reasons, we believe that a single quality rating for a sequence is insufficient in the stereoscopic scenario. We have thus developed an extended methodology, in which seven questions are answered by the subjects for each sequence presented. At the time of publishing our work, no standard subjective test methodology was available. A short time later, however, ITU-R approved Recommendation ITU-R BT.2021 "Subjective Methods for the Assessment of Stereoscopic 3DTV Systems" [18], in which visual (dis)comfort is also an important part of the scoring.

The different display technologies under study included a Liquid Crystal Display (LCD) with passive polarized glasses separation, a plasma display with synchronized active shutter glasses separation and, finally, a projection system with active shutter glasses separation. The content included blu-ray stereoscopic captures as the highest quality available, satellite broadcasting captures in side by side frame packing format, consumer grade camcorder shots and static images. It can be observed that the differences in quality of experience caused by different content are more significant than the display technologies employed.

The presented work was built upon by Fliegel et al. [7], who reproduced the test and performed and inter-laboratory comparison. A survey study by Urvoy et al. [59], focusing on visual fatigue and discomfort in 3D-TV, is mentioning our finding that there are differences in how different display technologies are influenced by ambient light. Liu et al. [26] focus on the impact of camera configurations on shooting quality in short distance conditions, extending our study with a set of objective parameters related to capture and quantifying their impact on QoE. Finally, Gutierrez et al. performed a similar comparison of stereoscopic display technologies [11]. Different from our study, also an autostereoscopic display was included and the set of QoE-related questions the observer had to answer was reduced to three.

3.2.3 Subjective Quality Assessment in Scalable Video for Measuring Impact Over Device Adaptation

Originally published as:
Lopez, J.; Slanina, M.; Arnaiz, L.; et al.: Subjective quality assessment in scalable video for measuring impact over device adaptation. In *2013 IEEE EUROCON*, July 2013, p. 162–169, DOI: 10.1109/EUROCON.2013.6624981.

Abstract:
Multiplicity of devices used for the reproduction of video is increasing nowadays. The characteristics of the displayable video are specific for each type of device. In order to avoid using separate source streams for the different devices and use optimal delivery bit rate, it is desirable that the encoded and transmitted original is identical for all the display devices and adapted in terms of video characteristics and bit rate with techniques such as Scalable Video Coding (SVC). In this paper, an analysis of perceived quality for different devices and different video characteristics is described, consisting in subjective tests with a group of observers and analysis of results. A collection of sequences were presented to the observers at different qualities, receiving a score on the Mean Opinion Score (MOS) scale (five grades from worst to the best quality). The devices considered included tablets and mobile phones,

and complemented previous studies developed with high definition televisions. The results of the tests are presented and conclusions are drawn.

Up to this point, our QoE-related work considered the subjective methodology selection and mainly content related issues. As mobile devices such as tablets and smartphones raised in popularity tremendously during the last five years, we could witness increasing multimedia consumption on the go. The paper *Subjective Quality Assessment in Scalable Video for Measuring Impact over Device Adaptation* [27] is focusing on the difference in perceived quality when the same content is viewed on different devices.

Due to bit rate and computational power / battery life constraints, it is usually required that different quality levels are available for streaming to a mobile device. Different quality levels can either be offered as separate streams to be offered to the client, or can be extracted from a full quality sequence on the fly if a scalable codec is used. This work uses the Scalable Video Coding (SVC) extension of the very popular H.264/AVC coding algorithm. The extension is described in Annex G of H.264/AVC [19].

When different bit rate (and different quality) video sequences need to be generated from a full quality video content, a decision is needed on what dimension of the sequence to change. One option is to decrease the spatial resolution of frames, i.e., reduce the frame dimensions in pixels. Another possibility is to decrease the temporal resolution, i.e., decrease the number of frames per second. Finally, the compression ratio of the encoder can be changed, at the expense of fine details being removed and compression artifact introduced. Motivated by these facts, the paper focuses on the following question:

- How do the different types of scalability (spatial, temporal, quality) affect the human perception? Is there a difference among various mobile display devices?

In order to answer this question, we have performed a subjective study using five source video sequences. We varied their resolution between 1280×720 and 176×144 pixels, their frame rate between 50 and 6.25 frames per second and changed the compression ratio through H.264/AVC quantization parameter adjustment between 20 and 45. We have found that the impact on perceived QoE is different on smartphones and tablets: Quite naturally, due to smaller screens, lower spatial resolution is better tolerated on smarphones. Decreased temporal resolution is, surprisingly, rated higher on tablets than on smartphones. Finally, coding quality reduction is perceived more strongly on tablets. An interesting finding is that for high coding quality (low Quantization Parameter value), smartphones show a saturation effect, i.e., coding becomes transparent much earlier than on tablets.

The work has become a reference for two different papers. Abis et al. [1] extend the experiment with the possibility of content quality change during the viewing session and also the possibility that the observer changes the device during the session, for example, moves from a TV to a tablet. The findings of our work have also been included in book chapter by Cheon et al., with an overview on quality assessment on mobile videos [5].

3.3 Scalable and Adaptive Video Streaming

As discussed in Subsection 3.2.3, different types of end devices to which multimedia is streamed have become a reality. These devices reportedly contribute to the user's Quality of Experience as well as content type, content properties and even context. Further, different devices come along with different requirements on the content properties as also discussed in Subsection 3.2.3.

So far, we have only considered the impact of video sequence parameters on the perceived quality on different devices. We didn't explicitly take bit rate into account, however bit rate is definitely the most important parameter when it comes to video or multimedia delivery in general. Having the possibility of selecting the properties and bit rate of the media being transmitted is extremely important in the usual case there is a limited or even varying bit rate available in the link between the media source (server or transmitter) and sink (client or receiver). Basically, such adjustment can be done in two ways:

- *Using a scalable video codec.* The scalable encoder uses a pre-defined hierarchy, so that lower resolution (spatial, temporal, quality) representations can be extracted and carried at a decreased bit rate. Scalable encoding, in principle, requires overhead information to provide such functionality, i.e., there is a certain bit rate increase traded off for flexibility.

- *Providing several copies of the same content but different quality and giving the client the freedom to select the most appropriate.* Such approach is obviously more suitable for streaming than for broadcasting services. The simplest approach is that the client selects the representation most suitable to its capabilities (e.g. spatial resolution) and typical available bit rate. A more sophisticated approach is to allow the client to switch among the different representations during the streaming session.

The two approaches target the same goal: tackle the heterogeneous environment of different client devices and support different transmission bitrates including rate variations. As further discussed in [40], there are advantages and disadvantages of both techniques.

In our work, we have at first [52] targeted scalable encoding techniques due to the very broad applicability – in principle, the lower resolution representation can be extracted from a scalable bit stream at any point during transmission, be it at the sender (server / transmitter), the transmission path or the receiver (client). As we found that in real world setup it is very difficult to achieve low overhead with SVC, our further research turned to the alternative approach described as HTTP Adaptive Streaming (HAS) [41], [40].

The content, results and impact of relevant published papers will be discussed in Subsections 3.3.1, 3.3.2 and 3.3.3, respectively.

3.3.1 Rate Distortion Performance of H.264/SVC in Full HD with Constant Frame Rate and High Granularity

Originally published as:
Slanina, M.; Ries, M.; Vehkapera, J.: Rate distortion performance of H.264/SVC

in full HD with constant frame rate and high granularity. In *The Eighth International Conference on Digital Telecommunications (ICDT)*, Venice, 2013, ISBN 978-1-61208-030-7, p. 7–13.

Abstract:
In this paper, we provide an analysis of performance of the scalable extension of the H.264/AVC video codec. We assume a fixed display scenario at full HD resolution with a constant frame rate. The encoded bit stream consists of one Coarse Grain Scalability (CGS) and one Medium Grain Scalability (MGS) layer with three sublayers, allowing for creating four quality layers from the complete bit stream. Hierarchical coding is employed, which results in dyadic decomposition of temporal layers. In order to increase the granularity of quality layers, the packets are selectively dropped from appropriate quality and temporal layers. We provide a performance analysis of such approach, compare the rate distortion performance to a mainstream H.264 encoder and analyze the composition of the bit stream at the considered operation points. Our findings show that quality enhancement of temporal layers has different effect on the overall performance depending on which temporal layer is enhanced.

Our work published in the paper *Rate Distortion Performance of H.264/SVC in Full HD with Constant Frame Rate and High Granularity* [52] is presenting a substream extraction scheme for increased granularity, i.e., offering a higher number of bit rate / quality level combinations than explicitly supported at the time of encoding.

The aim of the paper is two-fold:

- Verify the compression efficiency of the codec in a real world setup including high definition resolution, currently widely used in both television broadcasting and online streaming;
- Increase the granularity of substream extraction, i.e., the number of different quality levels that can be gained from the complete bit stream. The granularity increase should be achieved in the substream extraction process only.

The overview papers on H.264/SVC (e.g. [39]) use low resolution video content to test the codec principles and performance. Based on such analyses, it is widely understood that the bit rate overhead of H.264/SVC is adding approximately 10% on top of the bit rate required for non-scalable coding using H.264. In our test scenario, however, we have found that the bit rate overhead largely depends on the configuration of the encoder and overhead as low as 10% cannot easily be achieved in the high definition scenario. Then, non-scalable H.264/AVC is clearly significantly better in terms of rate-distortion efficiency.

The performance of scalable encoders in general depends on the encoder setup and hierarchy of different levels employed. As long as the layer dependencies are preserved, however, there is a relatively high freedom in the substream extraction. In our work we propose and analyze a high granularity substream extraction approach, in which quality refinement data from different temporal layers are discarded.

Zakerinasab et al. include our findings in their paper [69] and provide a deeper analysis of the optimal layer hierarchy for full high definition video streaming.

3.3.2 "To Pool or Not To Pool": A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming

Originally published as:
Seufert, M.; Slanina, M.; Egger, S.; et al.: "To pool or not to pool": A comparison of temporal pooling methods for HTTP adaptive video streaming. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013, p. 52–57, DOI: 10.1109/QoMEX.2013.6603210.

Abstract:
Current objective video quality metrics typically estimate video quality for short video sequences (10 to 15 sec) of constant quality. However, customers of video services usually watch longer sequences of videos which are more and more delivered via adaptive streaming methods such as HTTP adaptive streaming (HAS). A viewing session in such a setting contains several different video qualities over time. In order to express this in an overall score for the whole viewing session, several temporal pooling methods have been proposed in the related work. Within this paper, we set out to compare the performance of different temporal pooling methods for the prediction of Quality of Experience (QoE) for HTTP video streams with varying qualities. We perform this comparison based on ground truth rating data gathered in a crowdsourcing study in the context of the NGMN P-SERQU project. As input data for the models, we use objective video quality metrics such as PSNR, SSIM but also very basic inputs such as the bitrate of the clips only. Our results show that certain pooling methods perform clearly better than others. These results can help in identifying well performing temporal pooling methods in the context of HAS.

As already mentioned in Section 3.3, due to the findings of [52], the focus of our further work turned to the alternative variable bit rate approach of HTTP Adaptive Streaming. In the paper *"To Pool or Not To Pool": A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming* [41] we exploit a per-frame objective video quality metric for subjective quality estimation in the HAS context.

HAS is a simple yet efficient technique capable of adjusting the bit rate of the streaming media to the bit rate available in the network path from the server to the client. Depending on the implementation of the technique, the bit rate of the content transmitted may change every 2 or more seconds of video playback. An obvious result is that with the fluctuating bit rate of the compressed bit stream, the visual quality of the video played back fluctuates as well. Such behavior is, naturally, not reflected in the so-called per-frame objective metrics, which calculate a quality value for each video frame. How do we create a single quality indicator for a whole sequence then? The answer lies in temporal pooling, a technique used to combine a number of quality values into a single quality indicator. In our paper, we aim to answer the following question:

- Among the temporal pooling algorithms available, which is the most suitable for quality measurement in HAS?

As the performance indicator, we use the Pearson Linear Correlation Coefficient (PLCC) and Spearmant Rank Order Count Coefficient (SROCC) evaluated on the pooled values and a Mean Opinion Scores (MOS) achieved in a subjective experiment. In total, 13 temporal pooling algorithms were implemented and analyzed. Surprisingly, for total sequence length in the order of minutes, none of the sophisticated pooling algorithms performed significantly better than a simple mean of values.

The impact of the work is obvious in a number of citing papers. Miller et al. in [29] perform an evaluation of different adaptive streaming clients in a simulated wireless local area network using the achieved *average* video bit rate as one of the benchmark criteria. In [28] the authors consider QoE in adaptive streaming context a challenging task without deeper integration in their model. Marie-Neige Garcia et al. present an overview of video quality models for adaptive streaming in [9], where our model is identified as a bridge between existing modes for quality measurement in short sequences and adaptive streaming content with varying quality. Sackl et al. [38, 37] study the influence of network quality changes on QoE in a different context – Web QoE. Interestingly, they found that frequency of throughput alteration had no significant effect of subjective quality ratings. Also no differences in quality rating behavior between increasing and decreasing quality profiles was found, which confirms validity of our results even in a different application context.

Out of the available temporal pooling methods, a mean of frame scores was used based on our recommendation in [25, 57], while Horch et al. use average quality ratings per GOP to categorize quality fluctation patterns [13]. Sogaard et al. [53] use the mean and the standard deviation of per-frame objective metric scores as an input to a machine learning-based video quality metric.

In [8], the authors examine the validity of short-term quality models for long-term quality prediction, i.e. sequence length in the order of minutes. They find that short term quality models are good indicators of the long therm quality, independently of the pooling strategy. Staelens et al. present a continuous subjective quality of longer duration video sequences in the context of adaptive streaming in[56] to find that the amplitude of quality switches also has an impact on the overall QoE. Hossfeld et al. analyze the different QoE influence factors in HTTP adaptive streaming in [17]. With reference to our paper, the state that the quality level switching frequency is less important than the time spent on each layer in total.

Finally, Barkowsky et al. in [2] describe the principles to be employed when building a large and reliable database for development of enhanced video quality models. Our work is cited with respect to QoE degradation in adaptive streaming due to video stalling.

3.3.3 A Survey on Quality of Experience of HTTP Adaptive Streaming

Originally published as:
Seufert, M.; Egger, S.; Slanina, M.; et al.: A survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, 2015: p. 469–492, ISSN 1553-877X, DOI: 10.1109/COMST.2014.2360940.

Abstract:
Changing network conditions pose severe problems to video streaming in the Internet. HTTP adaptive streaming (HAS) is a technology employed by numerous video services which relieves these issues by adapting the video

> to the current network conditions. It enables service providers to improve resource utilization and Quality of Experience (QoE) by incorporating information from different layers in order to deliver and adapt a video in its best possible quality. Thereby, it allows to take into account end user device capabilities, available video quality levels, current network conditions, and current server load. For end users, the major benefits of HAS compared to classical HTTP video streaming are reduced interruptions of the video playback and higher bandwidth utilization, which both generally result in a higher QoE.
>
> Adaptation is possible by changing the frame rate, resolution, or quantization of the video, which can be done with various adaptation strategies and related client- and server-side actions. The technical development of HAS, existing open standardized solutions, but also proprietary solutions are reviewed in this article as fundament to derive the QoE influence factors which emerge as a result of adaptation.
>
> The main contribution is a comprehensive survey of QoE related works from human computer interaction and networking domains which are structured according to the QoE impact of video adaptation. To be more precise, subjective studies which cover QoE aspects of adaptation dimensions and strategies are revisited. As a result, QoE influence factors of HAS and corresponding QoE models are identified, but also open issues and conflicting results are discussed. Furthermore, technical influence factors, which are often ignored in the context of HAS, affect perceptual QoE influence factors and are consequently analyzed.
>
> This survey gives the reader an overview of the current state of the art and recent developments. At the same time it targets networking researchers who develop new solutions for HTTP video streaming or assess video streaming from a user centric point of view. Therefore, the article is a major step towards truly improving HAS.

Although the popularity of HAS has grown considerably in recent years, a thorough overview of the technical solutions, implementation possibilities and QoE implications was not available. One attempt for such thorough overview is our survey paper *A Survey on Quality of Experience of HTTP Adaptive Streaming* [40].

In our survey, we describe the principle of HAS and the different technical solutions available. Further, we analyze the different influence factors on QoE in HAS and compare the importance of different parameters based on a literature study. We also analyze the network behavior in case HAS is deployed and describe the scenarios in which client fairness is threatened. The papers mentioned in the following paragraphs use our work as reference.

Wamser et al. in [60] introduce a YouTube performance monitoring application for client-side measurement of video streaming on mobile devices supporting HAS. Zhou et al. propose an objective metric to assess QoE in HAS in order to enable QoE-aware resource allocation in a wireless local area network [70]. Sogaard et al. employ machine learning for the desing of an objective quality metric suitable for HAS in [53].

In [4], Casas et al. focus on the bit rate requirements of five popular applications in mobile devices, including YouTube with HAS. Timmerer et al. reference our study with respect to client adaptation logic strategies in HAS [58]. Hossfeld et al. in [12] performed a statistical evaluation of different downloading strategies taking network conditions into account. In [31], Ojanpera et al. target network optimization applicable when HAS is used in heterogeneous networks using a cognitive network management system while Bouten et al. use in-network quality optimization agents to monitor available throughput in [3]. Rubin et al., in contrast, consider resource allocation when HAS is used in a LTE network [36]. In[62], Wu et al. define a network-based video stalling detection approach for HAS while Petrangeli et al. try to proactively avoid freezes using dynamic prioritization of HAS streams in [32]. In [16], Hossfeld et al. describe a scenario using also out-of-band information (such as social network trend analytics) to help further optimize QoE. Devlic et al. [6] provide a QoE-aware optimization approach for video delivery where the content properties are adjusted for a specific display, constrained by the available transmission bit rate or storage space. Staelens et al. evaluate QoE when camera switching is employed in a HAS-based transmission of sports content in [55]. Finally, Hossfeld et al. in [15] use classification of users based on user profiles in order to create a queuing model for improving overall QoE in a network.

A survey on QoE in Video-on-Demand services, including HAS, has been published by Juluri et al. in [20].

Although the survey was published in the beginning of 2015, the listed citations are already available at the time of finishing this thesis (August 2015).

3.4 Research Work Beyond This Thesis

This thesis includes the papers which are the most important outputs of my research work. During my academic career, I have also been involved in other research work and authored or co-authored papers, which do not form an integral part of this thesis. These are several other papers belonging to the thesis topics, namely:

- More work on the no-reference metric design has been published in [49, 50]. The paper [67] presents a freely available no-reference video quality measurement tool. An overview of different approaches to video image quality measurement has been published in a book chapter [24].
- Preliminary results of our work the temporal aspect of subjective quality test sessions hav been published in [45]. In [34] we compare the results achieved in a subjective experiment, attained with or without reference in the session.
- In our further work on HAS, we review the available implementations of the new generation video coding standard in [66]. Then, in [68], we analyze the impact of adaptive streaming on QoE in case common high definition streaming is extended to ultra high definition.

Apart from the included topics, I have been involved in research work on modeling and measurement of wireless communication systems and their coexistence. In [23], we focus on modeling of indoor channel in two frequency bands, 0.9 GHz and 2.5 GHz. In [44], we focus on the relationship between the parameters of the LTE physical layer and the practically achievable throughput, while in [33], the performance of LTE and DVB-T2 in a shared frequency band is analyzed.

3.5 Bibliography

[1] Abis, N.; Floris, A.; Argyropoulos, S.; et al.: I have to switch the terminal: evaluating the impact on video quality perception. In *IEEE International Conference on Communications (ICC)*, London (UK), 2015, p. 1–6.

[2] Barkowsky, M.; Masala, E.; Wallendael, G. V.; et al.: Objective video quality assessment — Towards large scale video database enhanced model development. *IEICE Transactions on Communications*, vol. E98.B, no. 1, 2015: p. 2–11, DOI:10.1587/transcom.E98.B.2.

[3] Bouten, N.; Schmidt, R. d. O.; Famaey, J.; et al.: QoE-driven in-network optimization for adaptive video streaming based on packet sampling measurements. *Computer Networks*, vol. 81, 2015: p. 96–115, ISSN 1389-1286, DOI:10.1016/j.comnet.2015.02.007.

[4] Casas, P.; Schatz, R.; Wamser, F.; et al.: Exploring QoE in cellular networks: How much bandwidth do you need for popular smartphone apps? In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, AllThingsCellular '15, New York, NY, USA: ACM, 2015, ISBN 9781450335386, p. 13–18, DOI:10.1145/2785971.2785978.

[5] Cheon, M.; Kim, S.-J.; Chae, C.-B.; et al.: Quality assessment of mobile videos. In *Visual Signal Quality Assessment*, edited by C. Deng; L. Ma; W. Lin; K. N. Ngan, Springer International Publishing, 2015, ISBN 9783319103679 9783319103686, p. 99–127.

[6] Devlic, A.; Kamaraju, P.; Lungaro, P.; et al.: QoE-aware optimization for video delivery and storage. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, June 2015, p. 1–10, DOI:10.1109/WoWMoM.2015.7158122.

[7] Fliegel, K.; Vítek, S.; Jindra, T.; et al.: Comparison of stereoscopic technologies in various configurations. In *Applications of Digital Image Processing XXXV*, vol. 8499, 2012, p. 849929–849929–9, DOI:10.1117/12.929237.

[8] Garcia, M.; Robitza, W.; Raake, A.: On the accuracy of short-term quality models for long-term quality prediction. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, p. 1–6, DOI:10.1109/QoMEX.2015.7148123.

[9] Garcia, M.-N.; Argyropoulos, S.; Staelens, N.; et al.: Video streaming. In *Quality of Experience*, edited by S. Möller; A. Raake, T-Labs Series in Telecommunication Services, Springer International Publishing, 2014, ISBN 9783319026800 9783319026817, p. 277–297.

[10] Guntuku, S.; Scott, M.; Yang, H.; et al.: The CP-QAE-I: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, p. 1–7, DOI:10.1109/QoMEX.2015.7148130.

[11] Gutierrez, J.; Jaureguizar, F.; Garcia, N.: Subjective comparison of consumer television technologies for 3D visualization. *Journal of Display Technology*, vol. PP, no. 99, 2015: p. 1–1, ISSN 1551-319X, DOI:10.1109/JDT.2015.2448758.

[12] Hoßfeld, T.; Seufert, M.; Sieber, C.; et al.: Identifying QoE optimal adaptation of HTTP adaptive streaming based on subjective studies. *Computer Networks*, vol. 81, 2015: p. 320–332, ISSN 1389-1286, DOI:10.1016/j.comnet.2015.02.015.

[13] Horch, C.; Habigt, J.; Keimel, C.; et al.: Evaluation of video quality fluctuations using pattern categorisation. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2014, p. 117–122, DOI:10.1109/QoMEX.2014.6982306.

[14] Horch, C.; Keimel, C.; Habigt, J.; et al.: Length-independent refinement of video quality metrics based on multiway data analysis. In *2013 20th IEEE International Conference on Image Processing (ICIP)*, September 2013, p. 44–48, DOI:10.1109/ICIP.2013.6738010.

[15] Hosfeld, T.; Moldovan, C.; Schwartz, C.: To each according to his needs: Dimensioning video buffer for specific user profiles and behavior. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, p. 1249–1254, DOI:10.1109/INM.2015.7140476.

[16] Hosfeld, T.; Skorin-Kapov, L.; Haddad, Y.; et al.: Can context monitoring improve QoE? A case study of video flash crowds in the internet of services. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, p. 1274–1277, DOI:10.1109/INM.2015.7140480.

[17] Hossfeld, T.; Seufert, M.; Sieber, C.; et al.: Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2014, p. 111–116, DOI:10.1109/QoMEX.2014.6982305.

[18] ITU: Subjective methods for the assessment of stereoscopic 3DTV systems. ITU-R Recommendation BT.2021. 2012.
URL `http://www.itu.int/rec/R-REC-BT.2021-1-201502-I`

[19] ITU: Advanced Video Coding for Generic Audiovisual Services. ITU-T Recommendation H.264. 2014.
URL `https://www.itu.int/rec/T-REC-H.264`

[20] Juluri, P.; Tamarapalli, V.; Medhi, D.: Measurement of Quality of Experience of Video-on-Demand services: A survey. *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, 2015: p. 1–1, ISSN 1553-877X, DOI:10.1109/COMST.2015.2401424.

[21] Keimel, C.; Habigt, J.; Klimpke, M.; et al.: Design of no-reference video quality metrics with multiway partial least squares regression. In *2011 Third International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2011, p. 49–54, DOI:10.1109/QoMEX.2011.6065711.

[22] Keimel, C.; Klimpke, M.; Habigt, J.; et al.: No-reference video quality metric for HDTV based on H.264/AVC bitstream features. In *2011 18th IEEE International Conference on Image Processing (ICIP)*, September 2011, p. 3325–3328, DOI:10.1109/ICIP.2011.6116383.

[23] Klozar, L.; Prokopec, J.; Hanus, S.; et al.: Indoor channel modeling based on experience with outdoor urban measurement #x2014; Multislope modeling. In *2011 IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems (COMCAS)*, 2011, p. 1–4, DOI:10.1109/COMCAS.2011.6105905.

[24] Kratochvil, T.; Slanina, M.: Digital Video Image Quality. In *Digital Video*, edited by F. De Rango, InTech, February 2010, ISBN 9789537619701.
URL `http://www.intechopen.com/books/digital-video/digital-video-image-quality`

[25] Lin, J.; Wu, C.-H.; Katsavounidis, I.; et al.: EVQA: An ensemble-learning-based video quality assessment index. In *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, June 2015, p. 1–6, DOI:10.1109/ICMEW.2015.7169760.

[26] Liu, Y.; Yang, J.; Chu, R.: Objective evaluation criteria for shooting quality of stereo cameras over short distance. *Radioengineering*, vol. 24, no. 1, April 2015: p. 305–313, ISSN 1210-2512, DOI:10.13164/re.2015.0305, wOS:000353174400039.

[27] Lopez, J.; Slanina, M.; Arnaiz, L.; et al.: Subjective quality assessment in scalable video for measuring impact over device adaptation. In *2013 IEEE EUROCON*, July 2013, p. 162–169, DOI:10.1109/EUROCON.2013.6624981.

[28] Miller, K.; Bethanabhotla, D.; Caire, G.; et al.: A control-theoretic approach to adaptive video streaming in dense wireless networks. *IEEE Transactions on Multimedia*, vol. 17, no. 8, 2015: p. 1309–1322, ISSN 1520-9210, DOI:10.1109/TMM.2015.2441002.

[29] Miller, K.; Corda, N.; Argyropoulos, S.; et al.: Optimal adaptation trajectories for block-request adaptive video streaming. In *Packet Video Workshop (PV), 2013 20th International*, 2013, p. 1–8, DOI:10.1109/PV.2013.6691452.

[30] Montero, P.; Polak, L.; Taibo, J.; et al.: Subjective quality assessment of the impact of buffer size in fine-grain parallel video encoding. *Radioengineering*, vol. 21, no. 4, 2012: p. 1226–1235, ISSN 1210-2512.

[31] Ojanperä, T.; Luoto, M.; Majanen, M.; et al.: Cognitive network management framework and approach for video streaming optimization in heterogeneous networks. *Wireless Personal Communications*, March 2015: p. 1–31, ISSN 0929-6212, 1572-834X, DOI:10.1007/s11277-015-2519-7.

[32] Petrangeli, S.; Wauters, T.; Huysegems, R.; et al.: Network-based dynamic prioritization of HTTP adaptive streams to avoid video freezes. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, p. 1242–1248, DOI:10.1109/INM.2015.7140475.

[33] Polak, L.; Klozar, L.; Kaller, O.; et al.: Study of coexistence between indoor LTE femtocell and outdoor-to-indoor DVB-T2-Lite reception in a shared frequency band. *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, April 2015: p. 114, ISSN 1687-1499, DOI:10.1186/s13638-015-0338-x.
URL `http://jwcn.eurasipjournals.com/content/2015/1/114/abstract`

[34] Polak, L.; Slanina, M.; Kratochvil, T.; et al.: A comparison of single and double stimulus subjective assessment of full HD video sequences. In *Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference*, April 2014, p. 1–4, DOI:10.1109/Radioelek.2014.6828436.

[35] Ries, M.; Slanina, M.; Garcia, D.: Reference free SSIM estimation for full HD video content. In *Radioelektronika (RADIOELEKTRONIKA), 2011 21st International Conference*, April 2011, p. 1–4, DOI:10.1109/RADIOELEK.2011.5936447.

[36] Rubin, I.; Colonnese, S.; Cuomo, F.; et al.: Mobile HTTP-based streaming using flexible LTE base station control. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, June 2015, p. 1–9, DOI:10.1109/WoWMoM.2015.7158123.

[37] Sackl, A.; Casas, P.; Schatz, R.; et al.: Quantifying the impact of network bandwidth fluctuations and outages on Web QoE. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, p. 1–6, DOI:10.1109/QoMEX.2015.7148078.

[38] Sackl, A.; Egger, S.; Schatz, R.: The influence of network quality fluctuations on Web QoE. In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2014, p. 123–128, DOI:10.1109/QoMEX.2014.6982307.

[39] Schwarz, H.; Marpe, D.; Wiegand, T.: Overview of the Scalable H.264/MPEG4-AVC Extension. In *2006 IEEE International Conference on Image Processing*, 2006, p. 161–164, DOI:10.1109/ICIP.2006.312374.

[40] Seufert, M.; Egger, S.; Slanina, M.; et al.: A survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, 2015: p. 469–492, ISSN 1553-877X, DOI:10.1109/COMST.2014.2360940.

[41] Seufert, M.; Slanina, M.; Egger, S.; et al.: "To pool or not to pool": A comparison of temporal pooling methods for HTTP adaptive video streaming. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013, p. 52–57, DOI:10.1109/QoMEX.2013.6603210.

[42] Shahid, M.; Rossholm, A.; Lövström, B.; et al.: No-reference image and video quality assessment: a classification and review of recent approaches. *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, August 2014: p. 40, ISSN 1687-5281, DOI:10.1186/1687-5281-2014-40.

[43] Shi, Z.; Chen, P.; Feng, C.; et al.: Research on quality assessment metric based on H.264/AVC bitstream. In *2012 International Conference on Anti-Counterfeiting, Security and Identification (ASID)*, 2012, p. 1–5, DOI:10.1109/ICASID.2012.6325335.

[44] Slanina, M.; Klozar, L.; Hanus, S.: Practical measurement of data throughput in LTE network depending on physical layer parameters. In *Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference*, April 2014, p. 1–4, DOI:10.1109/Radioelek.2014.6828484.

[45] Slanina, M.; Kratochvil, T.; Polak, L.; et al.: Temporal aspects of scoring in the user based quality evaluation of HD video. In *2011 34th International Conference on Telecommunications and Signal Processing (TSP)*, 2011, p. 598–601, DOI:10.1109/TSP.2011.6043659.

[46] Slanina, M.; Kratochvil, T.; Polak, L.; et al.: Analysis of temporal effects in quality assessment of high definition video. *Radioengineering*, vol. 21, no. 1, 2012.

[47] Slanina, M.; Kratochvíl, T.; Ricny, V.; et al.: Testing QoE in different 3D HDTV technologies. *Radioengineering*, vol. 21, no. 1, 2012.

[48] Slanina, M.; Ricny, V.: Estimating H.264/AVC video PSNR without reference using the artificial neural network approach. In *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, Porto, 2008, p. 244–250.

[49] Slanina, M.; Ricny, V.: Estimating H.264/AVC video PSNR without reference using the artificial neural network approach. In *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, Porto, 2008, p. 244–250.

[50] Slanina, M.; Ricny, V.: Estimating PSNR in High Definition H.264/AVC Video Sequences Using Artificial Neural Networks. *Radioengineering*, vol. 17, no. 3, September 2008: p. 103–108, ISSN 1210-2512, wOS:000261572600018.

[51] Slanina, M.; Ricny, V.; Forchheimer, R.: A novel metric for H.264/AVC no-reference quality assessment. In *14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, June 2007, p. 114–117, DOI:10.1109/IWSSIP.2007.4381166.

[52] Slanina, M.; Ries, M.; Vehkapera, J.: Rate distortion performance of H.264/SVC in full HD with constant frame rate and high granularity. In *The Eighth International Conference on Digital Telecommunications (ICDT)*, Venice, 2013, ISBN 978-1-61208-030-7, p. 7–13.

[53] Sogaard, J.; Forchhammer, S.; Brunnstrom, K.: Quality assessment of adaptive bitrate videos using image metrics and machine learning. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, p. 1–2, DOI:10.1109/QoMEX.2015.7148105.

[54] Song, R.; Jiang, X.; Shi, H.: Objective quality assessment of HD video based on compressed domain. In *International Conference on Graphic and Image Processing (ICGIP)*, vol. 8285, 2011, p. 828531–828531–6, DOI:10.1117/12.913742.

[55] Staelens, N.; Coppens, P.; Van Kets, N.; et al.: On the impact of video stalling and video quality in the case of camera switching during adaptive streaming of sports content. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, p. 1–6, DOI:10.1109/QoMEX.2015.7148102.

[56] Staelens, N.; De Meulenaere, J.; Claeys, M.; et al.: Subjective quality assessment of longer duration video sequences delivered over HTTP adaptive streaming to tablet devices. *IEEE Transactions on Broadcasting*, vol. 60, no. 4, 2014: p. 707–714, ISSN 0018-9316, DOI:10.1109/TBC.2014.2359255.

[57] Suárez, F. J.; García, A.; Granda, J. C.; et al.: Assessing the QoE in video services over lossy networks. *Journal of Network and Systems Management*, February 2015: p. 1–24, ISSN 1064-7570, 1573-7705, DOI:10.1007/s10922-015-9343-y.

[58] Timmerer, C.; Weinberger, D.; Smole, M.; et al.: Live transcoding and streaming-as-a-service with MPEG-DASH. In *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, June 2015, p. 1–4, DOI:10.1109/ICMEW.2015.7169746.

[59] Urvoy, M.; Barkowsky, M.; Le Callet, P.: How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *Annals of Telecommunications-Annales Des Telecommunications*, vol. 68, no. 11-12, December 2013: p. 641–655, ISSN 0003-4347, DOI:10.1007/s12243-013-0394-3, wOS:000327452400006.

[60] Wamser, F.; Seufert, M.; Casas, P.; et al.: YoMoApp: A tool for analyzing QoE of YouTube HTTP adaptive streaming in mobile networks. In *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, p. 239–243, DOI:10.1109/EuCNC.2015.7194076.

[61] Wang, Z.; Wang, W.; Wan, Z.; et al.: No-reference hybrid video quality assessment based on partial least squares regression. *Multimedia Tools and Applications*, July 2014: p. 1–14, ISSN 1380-7501, 1573-7721, DOI:10.1007/s11042-014-2166-0.

[62] Wu, T.; Huysegems, R.; Bostoen, T.: Scalable network-based video-freeze detection for HTTP adaptive streaming. In *IEEE/ACM International Symposium on Quality of Sevice*, Portland, 2015, p. 1–10.

[63] Wulf, S.; Zolzer, U.: Full-reference video quality assessment on high-definition video content. In *2012 6th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2012, p. 1–10, DOI:10.1109/ICSPCS.2012.6507948.

[64] Xu, J.; Jiang, X.: No reference estimation of the coding PSNR for 4K-UHD videos. In *International Conference on Cyberspace Technology (CCT 2014)*, 2014, p. 1–4, DOI:10.1049/cp.2014.1369.

[65] Xu, J.; Jiang, X.: No reference video quality metric for 4K-UHD videos based on H.264/AVC bitstream features. In *International Conference on Computer Information Systems and Industrial Applications (CISIA 2015)*, Atlantis Press, 2015, ISBN 9789462520721, p. 83–86, DOI:10.2991/cisia-15.2015.22.

[66] Zach, O.; Slanina, M.: A comparison of H.265/HEVC implementations. In *ELMAR (ELMAR), 2014 56th International Symposium*, September 2014, p. 1–4, DOI:10.1109/ELMAR.2014.6923337.

[67] Zach, O.; Slanina, M.: A Matlab-Based Tool for Video Quality Evaluation without Reference. *Radioengineering*, vol. 23, no. 1, April 2014: p. 405–411, ISSN 1210-2512, wOS:000334729600015.

[68] Zach, O.; Slanina, M.: A comparison of HAS behavior in the high definition and ultra high definition context. In *Radioelektronika (RADIOELEKTRONIKA), 2015 25th International Conference*, April 2015, p. 305–308, DOI:10.1109/RADIOELEK.2015.7129041.

[69] Zakerinasab, M.; Wang, M.: An anatomy of SVC for full HD video streaming. In *2014 IEEE 22nd International Symposium on Modelling, Analysis Simulation of Computer and Telecommunication Systems (MASCOTS)*, September 2014, p. 345–354, DOI:10.1109/MASCOTS.2014.50.

[70] Zhou, H.; Li, B.; Yang, M.; et al.: QoE-aware admission control and MAC layer parameter configuration algorithm in WLAN. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, p. 1066–1071, DOI:10.1109/WCNC.2015.7127617.

Abstract

The habilitation thesis is oriented in the field of performance assessment for multimedia systems enabling the emerging multimedia services. It is composed as commented proceedings of 9 author's most important research papers, published in peer reviewed international conferences or international journals in the years 2007 – 2015. The research papers are organized in three dominant topics: (1) Metrics for No-Reference Video Quality Measurement, (2) Quality of Experience in Various Viewing Contexts, and (3) Scalable and Adaptive Video Streaming.

Abstrakt

Tato habilitační práce je zaměřena na oblast hodnocení kvality pro multimediální systémy, které se stávají předpokladem pro příchod nových multimediálních služeb. Práce je pojata jako sborník devíti nejdůležitějších autorových publikací, zveřejněných ve sbornících recenzovaných konferenčních článků nebo v mezinárodních vědeckých časopisech v letech 2007 – 2015. Články jsou rozčleněny do tří dominantních témat: (1) Metriky pro měření kvality videa bez reference, (2) Kvalita zážitku v různém kontextu, (3) Škálovatelné a adaptivní streamování videa.