Multimed Tools Appl
DOI 10.1007/s11042-014-1978-2

Hybrid video quality prediction: reviewing video quality measurement for widening application scope

Marcus Barkowsky · Iñigo Sedano · Kjell Brunnström · Mikołaj Leszczuk · Nicolas Staelens

Abstract A tremendous number of objective video quality measurement algorithms have been developed during the last two decades. Most of them either measure a very limited aspect of the perceived video quality or they measure broad ranges of quality with limited prediction accuracy. This paper lists several perceptual artifacts that may be computationally measured in an isolated algorithm and some of the modeling approaches that have been proposed to predict the resulting quality from those algorithms. These algorithms usually have a very limited application scope but have been verified carefully. The paper continues with a review of some standardized and well-known video quality measurement algorithms that are meant for a wide range of applications, thus have a larger scope. Their individual artifacts prediction accuracy is usually lower but some of them were validated to perform sufficiently well for standardization. Several difficulties and shortcomings in developing a general purpose model with high prediction performance are identified such as a common objective quality scale or the behavior of individual indicators when confronted with stimuli that are out of their prediction scope. The paper concludes with a systematic framework approach to tackle the development of a hybrid video quality measurement in a joint research collaboration.

M. Barkowsky (✉)
LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597, Rue Christian Pauc, 44306 Nantes, France
e-mail: Marcus.Barkowsky@univ-nantes.fr

I. Sedano
TECNALIA, ICT - European Software Institute, Parque Tecnológico de Bizkaia, Edificio 202, 48170 Zamudio, Spain
e-mail: inigo.sedano@tecnalia.com

K. Brunnström
Department of Netlab, Acreo Swedish ICT AB (and Mid Sweden University), Stockholm, Sweden
e-mail: kjell.brunnstrom@acreo.se

M. Leszczuk (✉)
AGH University of Science and Technology, al. Mickiewicza 30, 30059 Kraków, Poland
e-mail: leszczuk@agh.edu.pl

N. Staelens
Department of Information Technology, Ghent University - iMinds, Ghent, Belgium
e-mail: nicolas.staelens@intec.ugent.be

Published online: 24 April 2014

Keywords Video quality assessment · Human visual system · Hybrid model development · Perceptual indicators · Quality of Experience

1 Introduction

Video quality assessment has been an important topic during the last decades, notably in the television broadcasting domain. The continuous spread of coverage and the transmission bit-rate increase in Internet connectivity at home led not only to the expected increase of IPTV services and consumption but also to an unexpected number of added-value services, both from commercial providers, such as downloadable video content and video-on-demand offers as well as customer-to-customer offers such as video sharing platforms.

Customer satisfaction, notably in terms of transmitted video quality, is the key to the success of these services. Topics which had limited application previously are gaining more interest now. One example in which automated measurement of perceived video quality plays an important role is the inter-channel bitrate allocation. It is well known that fast moving content such as sports requires higher bitrates than slow content, such as a single person presentation of the news in order to provide a comparable perceived quality. On satellite channels and in Digital Video Broadcasting (DVB), often only a limited number of content channels could be bundled, two to eight being a typical limitation. The combined bit-rate was fixed. Splitting the bit-rate unequally between the various programs was therefore limited to quantized values that did not vary significantly over time. This is different when the Internet is used as transmission channel. Each household with its DSL subscriber line may demand several video streams from different server locations. Head-end stations accumulate several subscriber lines and the combined traffic volume may be limited. In a spatially local area, several hundred video bit-streams may be streamed at the same time, accounting to and benefitting from a high temporal flexibility in individual bitrate allocation. In the ideal case, all subscribers will receive the desired Quality of Experience (QoE) for their service which is continuously monitored by automated measurements taking into consideration the transmitted bitstream features and the decoded video's quality instead of randomly dropping packets.

The objective measurement of QoE in the network or at the receiving side is the key aspect of this scenario. QoE is often used with ambiguous meaning. In this paper, we understand QoE in the way that has recently been defined as "*the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state*" [35].

This paper focuses on the prediction of one aspect of QoE: *perceived video quality*. It has important interactions with other aspects of QoE such as audio, environmental setup, or interactivity which are outside of the scope of this paper.

Objective video quality models can be divided into different categories depending on whether they use reference information or not, as well as if they are using bit-stream and/or video data. Historically, as most video quality models worked only on the video data, the terms Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) were used (and are still used) for describing the amount of information from the reference video. In this case, the reference video corresponds to a high quality uncompressed version of the video. FR models will then require access to the complete reference or original video. An RR model, on the other hand, will extract key information from both the reference and the degraded video, and compare these. As such, only a reduced amount of data needs to be transferred either to the sending point or the receiving point in order to compare the videos and predict perceived

quality. An NR model is working only on the degraded video and does not require a reference. They are also sometimes called zero reference.

As per the ITU standardization activities, the objective quality measurement methods have been classified into the following five main categories depending on the type of input data that is being used for quality assessment: media-layer models, parametric packet-layer models, parametric planning models, bitstream-layer models and hybrid models [13, 55]. Parametric *packet-layer models* inspect only the packet header such as IP-packets e.g. Real-time Transmission Protocol (RTP) or User Datagram Protocol (UDP), and from this information make predictions of the quality [25]. Such models are particularly interesting when video traffic is encrypted, or in the case of Digital Rights Management (DRM). *Bit-stream models* go further and also analyze the encoded (impaired) bit-stream itself and extract the information needed to make a quality prediction without fully decoding and processing the actual video [26]. *Hybrid models* use both video pixel information in combination with the bit-stream information eventually also doing a full decoding of the video payload.

Media-layer models use the speech or video signal to compute the Quality of Experience (QoE). These models do not require any information about the system under testing. *Parametric planning models* make use of quality planning parameters for networks and terminals to predict the QoE. As a result they require a priori knowledge about the system that is being tested. In [13] the *media-layer models* are classified into traditional point-based metrics (e.g., MSE and PSNR), Natural Visual Characteristics oriented metrics, and Perceptual (HVS) oriented metrics. The Natural Visual Characteristics metrics are further classified into Natural Visual Statistics and Natural Visual Features based methods. Similarly, the HVS methods are further classified into DCT domain, DWT domain and pixel domain models. The objective methods can also be classified in terms of their usability in the context of adaptive streaming solutions as out-of-service methods and in-service methods. In the out-of-service methods, no time constraints are imposed and the original sequence can be available.

We can generalize the concept of FR, RR and NR by also including packet header models, bit-stream models and hybrid models together with the pure video based models, providing a classification based on the amount of reference information used by the models, such as identifying a hybrid Bitstream-NR model.

The usefulness and applicability of these models are quite different. It has been considered that the most accurate models are the FR models; the VQEG phase II test did show that RR models can be at least as accurate [61]. The FR models and RR models can therefore be used for offline tuning of encoders and comparison between them. They can also be useful for generating training data for the development of NR models [11]. RR models are often argued to be an alternative for quality monitoring purposes in the network, but practically they are not. The reason is that the reference information that is processed at the sender side to generate the auxiliary RR information requires in most cases to be based on high quality, preferably uncompressed, versions of the videos. In practice, the network provider is not in possession of the same video as the content provider sends a coded video signal for distribution. This version of the video is often of quite low quality in itself and most RR models are not designed to compare towards a low quality reference.

Research on Hybrid-NR perceptual measurements faces two main categories of problems. The first category relates to technical details, such as capturing a transmitted video bit-stream, parsing the information from the current complex video compression standards, accessing the decoded video's pixel data, etc. The second category relates to the complexity of algorithmically modeling the Human Visual System (HVS). This comprises some prominent questions and fundamental research such as contrast sensitivity, spatial and motion masking, visual saliency, etc.

The first category can be tackled by centralizing the effort and managing technical tools for the research community in a collaborative effort. This has been recently established by the Joint Effort Group (JEG) of the Video Quality Experts Group, notably in their Hybrid project [52, 60]. Within JEG-Hybrid, a number of tools are made freely available for, amongst other, automatic monitoring and gathering of information at the video and network level. All information is gathered in structured XML files, which enable easy processing of the captured data. Furthermore, instead of having to parse the received encoded video bit-streams, JEG-Hybrid proposed the use of an XML-based file structure for representing the content of the (impaired) video bit-stream. These Hybrid Input XML files (HMIX files) contain information in a human readable format and enable quick and efficient processing of the video instead of having to write a complete parser. All software tools can be downloaded from VQEG's Tools and Subjective Labs Setup website [54].

The second category, the development of algorithmic prediction of the behavior of the HVS, is subject of this paper. In most publications, either an isolated aspect of the HVS is analyzed in detail, or a (complex) measurement algorithm is presented and verified on a video database annotated with subjective voting. The first approach lacks the verification on general content; the second approach misses to prove that the HVS has been modeled successfully.

A recent review of objective video quality measurement algorithms has been published in [13]. The authors investigated into a classification of the general methodology for predicting video quality in categories such as natural visual statistics or frequency domain HVS based. They also compared the performance of several models on a subjective database.

The approach of this paper is to first identify typical perceptual degradations that may be annotated by naive human observers, for example in Open Profiling approaches [53]. Publications tackling these isolated perceptual degradations will be referenced. In a second step, existing published (complex) models are analyzed with respect to their capability of predicting these perceptual artifacts.

The paper is limited in the sense that it cannot and will not exhaustively cover one or the other. It shall provide an incomplete list of algorithms to measure isolated degradations that were previously used in order to stimulate comparison. It shall further show which types of degradations measurement have previously been used in combined measurement algorithms. A continuous update of this list is expected to take place within the VQEG JEG group. The goal of this paper is to help in the identification of the artifacts and their measurement algorithms that may form the advantage of a Hybrid-NR measurement, providing access to both bit-stream measurement indicators and pixel based indicators.

The paper is organized as follows. In Section 2, the development of objective video quality prediction algorithms is introduced. For one of the identified components, the perceptual features and their algorithmic indicators existing methods are reviewed in Section 3. Complex algorithms for video quality measurement are reviewed in Section 4 with particular emphasis on their indicators. Section 5 proposes methods to identify the scope of the indicators and suggests optimization criteria. Currently ongoing collaborative efforts are documented in Section 6 before the paper is concluded in Section 7.

2 Development of objective video quality prediction

The structure of a video quality measurement algorithm is often a mixture of different concepts. The concepts originate from different methods to tackle the algorithmic prediction process.

The first method consists of mimicking the processing steps of the HVS. This is mostly applicable in the pixel domain. Typically, the reconstruction of the visual scene from the sender's side by the display device with all its limits is modeled first. The required steps include knowledge about the display device properties such as size, maximum illumination, and gamma curve, the position of the observer relative to the display device and environmental factors such as surrounding light sources that may lead to reflections from the screen and that may change the observer's pupil size. The technical background may be found in [42, 45, 47] and a practical example of the required steps is provided in [57]. Depending on the targeted precision of the model, the eye optics is respected, and biomedical modeling of retinal processes is considered. Further modeling of higher level processes may be taken into consideration provided that corresponding models exist. Most of these models are based on an explanatory theory of the outcome of several psychophysical experiments. The interested reader is referred to [8, 22]. Due to the complexity of the HVS and the close interaction with scene interpretation from memory, the processing usually only covers the first few steps of perception. This method is most efficient for degradations which are close to perception thresholds, normalizing the video input data to the human's detection threshold may also improve performance for supra-threshold, i.e. immediately visible degradations [40].

The second approach also takes into consideration isolated aspects that are recognizable by the human observer but does not aim at a constructive reproduction of the HVS's processing chain. Instead, psychophysical experiments are conducted in order to model the relationship to the quality prediction for a complex but isolated aspect. One example is the required frame-rate for video transmission. Psychophysical experiments on the contrast sensitivity threshold using sinusoidal stimuli have been conducted in [15]. The appearance for natural video sequences and more importantly the influence on perceived quality may be different. Therefore, several experiments have been conducted with natural video sequences in order to evaluate the quality degradation. This led to algorithms that may predict this particular perceptual degradation which depends also on the video resolution as will be explained later in Section 5. This approach is usually used when the artifacts are too complex for integrating them in the HVS's processing chain and require interpretation by human observers.

For its measurement, two different approaches exist to estimate the impact on the human observer: The first approach is that the required technical parameter is known a priori or can be extracted easily from the received data such as the spatial resolution. The second situation occurs when the parameter is not directly accessible but needs to be estimated either from bit-stream data or from the decoded video sequence. An example for this type of analysis would be blockiness and blurriness which can be either measured using a spatial frequency analysis of the decoded video or by estimation using the QP value and the distribution of the residual Discrete Cosine Transform (DCT) coefficients.

A later step for the development of a video quality measurement algorithm is to take into consideration the interaction of different artifacts, including the training on specifically designed subjective experiments that use natural content in high quality as reference and that perform typical processing steps for the planned application area of the objective quality prediction. The combinations of different artifacts that result from these processing steps are then evaluated in a subjective experiment to measure the impact of the complex combination of artifacts.

The typical development cycle of a measurement method is depicted in Fig. 1. The top of the diagram contains the subjective assessment experiments that are required for developing a perceptual measurement. On the left hand side, the isolated subjective experiments for individual quality degradations are shown with the three steps of isolating a psychophysical effect such as contrast sensitivity, designing appropriate stimuli, such as Gabor patterns, and

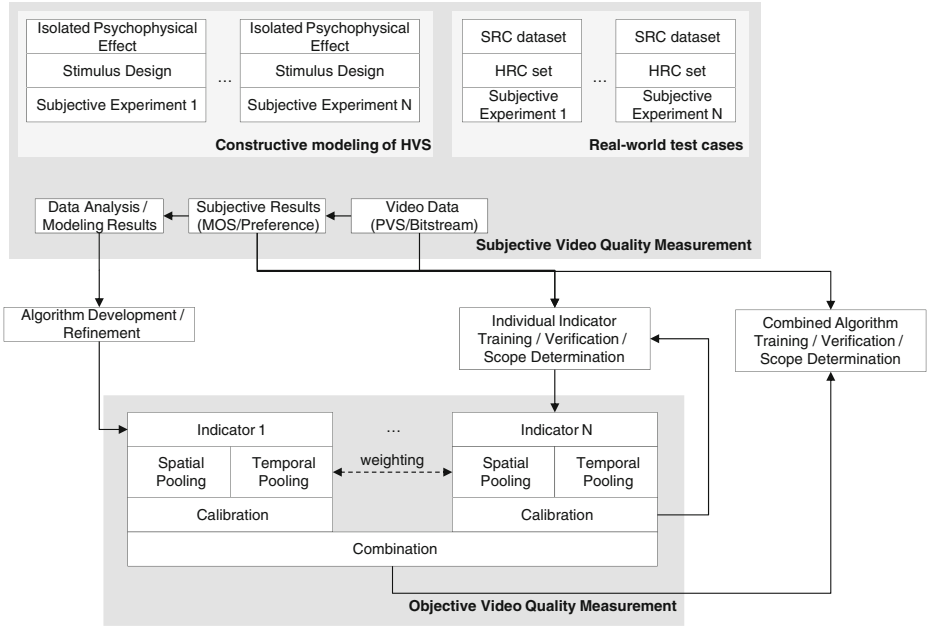

Fig. 1 Development steps for an objective video quality measurement algorithm

finally conducting the subjective experiment using appropriate setup and methodology. Commonly used subjective assessment methodologies have been standardized in International Telecommunications Union (ITU)—T Recommendations P.910 and P.911, and in ITU-R Rec. BT.500. On the right hand side, the creation of the datasets for combined artifact measurement is detailed as consisting of the selection of the Source Reference Sequences (SRCs), the choice of the Hypothetical Reference Circuits (HRCs) such as a specific video encoder and transmission chain, and finally the subjective experiment which eventually requires different setup and methodology.

Each of these experiments usually leads to three outcomes. The first and most immediate result is the video data that was shown to the observers, mostly in the form of processed video sequences (PVS) and eventually bit-stream data. The second result is the set of votes that were given by the observers of the same. Finally, the third outcome is the analysis of the votes with respect to the degradations presented in the video files that may be modeled and implemented in an algorithm to automatically predict the outcome of the subjective experiment. Often, verification is also performed showing the performance on the same data as was used for the development and training. Seldom, validation is performed, i.e. measuring performance by using a different data set, preferably created by a different process and/or research group.

The lower part of the diagram shows the processing performed in an algorithm, starting by individual indicators that may interact in terms of weighting during the spatial and temporal pooling stage. A typical example being visual saliency that may be employed for learning about the size of the most interesting object shown on the screen but which may also be used to weight degradations with information on relative spatial importance.

It should be noted that the conceptual diagram representation may differ from the implementation. In particular, the integration of required and eventually shared preprocessing steps

such as color space conversions into each indicator allows for easier modularization. In implementations, redundancy may be removed for speed gains. For example, caching structures in object oriented frameworks may be employed.

From the diagram, it becomes obvious that the prediction performance of a video quality measurement algorithm depends on the quality of its indicators and the combination of the indicators. This may be helped by repeated verification procedures during the development using subjectively evaluated datasets and rigorous evaluation of the performance of each indicator within the complete framework.

It is evident that the combination of a larger set of indicators and the training of their combinations poses stability issues with more complex training methods such as neural networks while simpler training methods may result in suboptimal prediction. For similar reasons, the prediction performance depends on the size, quality, and feature spread of the training and verification databases.

The literature currently lists, on the one hand, a large number of possible indicator algorithms and combination methods. On the other hand, there is a long list of complete algorithms for video quality prediction. The two following sections will review a few examples of each category.

3 Perceptual features and their algorithmic indicators

The term perceptual feature will be used in the following to identify and describe an isolated aspect of perception, explicitely or implicitely experienced by the human observer. The list of perceptual features mentioned in the literature is endless. The following list aims at listing important indicators on the one hand side and on demonstrating the variety of perceptual indicators on the other side. While only a short excerpt will be given in this paper, a longer and continuously updated list will be established and made available by the VQEG-JEG group on their homepage [60].

3.1 Overall still image based features

In most cases, video quality assessment algorithms analyze first individual images and then perform temporal pooling for determining the quality of the complete video sequence. The overall quality of this processing step may be identified.

Still Image Difference to Reference: For each individual image, most FR measurement algorithms estimate an individual quality value that corresponds to the perceived difference between the original and the degraded version. For RR and NR measurements, a difference to an optimal reference image is often predicted. Human observers may be asked to rate this difference in an image based double stimulus paired comparison experiment on a 7-point adjectival categorical judgement scale on selected images of the video sequence.

3.2 Classification of content related perception

Without taking into consideration the degraded image, several properties of the content itself may be taken into consideration. These factors play either an important role in the weighting of the degradation perception, such as masking effects, or they relate to a different kind of interpretation of the perceived content, for example when cartoons or drawings are evaluated.

3.2.1 Contrast sensitivity

It has been clear for a long time that the human visual system is differentially sensitive to patterns containing different spatio-temporal frequencies usually referred to contrast sensitivity, which is described by the contrast sensitivity function (CSF) [6]. A standard model for contrast detection based on the spatial CSF was developed by [63]. The CSF is different for achromatic and chromatic channels, where the achromatic CSF has higher sensitivity for high frequencies as compared to the chromatic, which has also been successfully utilized by standard colour format e.g. YUV 4:2:0 [46], where the achromatic channel Y has double resolution in each direction as compared to the chromatic channels U and V. In principle the spatio-temporal CSF is not space-time separable as pointed out by e.g. [33], but others have later found reasonable approximations of the CSF separable in space and time [12]. Also the contrast sensitivity can be model by the inspiration of the connection between the ganglion cells and the lateral geniculate nucleus (LGN), where the processing is divided in two channels, Parvo and Magno, where Parvo is loosely taking care of the high spatial frequencies and low temporal frequencies, whereas Magno is taking care of the low spatial frequencies and high temporal frequencies [1, 10].

3.2.2 Masking

The visibility of a pattern is affected by the pattern surrounding it. Legge and Foley defined masking as "any destructive or interference among transient stimuli that are closely coupled in space and time" [19, 36]. The effect is mostly negative, that is the effect is that the detectability of pattern is decreased. However, this is not always the case for some low contrast masker the detectability may actually increase. Masking has been exploited in several image discrimination models e.g. [1, 39, 64] and video discrimination models e.g. [9, 59, 66].

3.2.3 Spatial masking

The presence of a texture pattern in an image region may mask the visibility of similar patterns in its neighborhood. This has been studied with various psychophysical stimuli and led to different models [36]. A computational model for natural image content at typical luminance levels for broadcast TVs has been discussed in [44].

3.2.4 Temporal masking

Moving textured regions not only attract attention but also mask surrounding textures [20]. A study of a moving edge masking has been analyzed in [21]. A special case of temporal masking in video quality analysis is scene cuts, a detailed analysis may be found in [56].

3.3 Classification of spatial degradations

Technical constraints such as maximal allowed bit-rate allocation often introduce a perceptually complex pattern of degradations. The degradations have been isolated and categorized. This allows for studying the origin of each such degradation, for example in a video coding algorithm. It also allows for developing individual algorithms that measure each such degradation.

3.3.1 Blockiness

Blockiness or block artifacts are often caused by the use of lossy compression. This stems from independent transform coding of "NxN" pixel blocks (usually 4×4 or 8×8 pixels) in most of the currently used video coding algorithms including *H.261-H.265*, *MPEG-4 Part 2* and *Part 10* or *MPEG-2*. These algorithms use a quantization of the DCT coefficients for each block separately, which causes noise shaping that leads to coding artifacts in the form of discontinuities for coded block boundary [65]. Sudden color intensity changes are most evident in uniform areas of an image and are caused by the removal of the least significant coefficients of DCT.

Block artifacts can be calculated locally for each coding block. For example, the absolute difference in luminance pixels in pairs of adjacent pixels within a coding block and the pair of pixels of neighboring blocks may be calculated. Then the total value of differences within the blocks and between the blocks may be calculated and averaged over the entire frame [51].

3.3.2 Blurriness

Blur is mostly caused by the removal of DCT coefficients of high frequency or by introduction of loop-filters to counteract on blockiness, which both lead to low-pass filtering. This effect can be seen as a loss of detail in the image, reducing sharp edges and texture of objects. Moderate blur effects may occur due to loop-filters in current encoding standards or due to the combination of image patches from bi-directionally predicted coding-blocks. While these effects usually lead to perceived smoothness for luminance signals, the same effects on chrominance coding may lead to *smearing* on the edges of areas with contrasting color values.

Measurement of this artifact may be based on the cosine of the angle between perpendiculars to planes in adjacent pixels [51].

3.3.3 Geometric distortions

Geometric distortions may be caused by various types of image adaptations such as re-scaling due to aspect ratio conversion. For objective quality assessment, this artifact requires not only the measurement of the perceptual impact on quality but also a sophisticated adaptation to a non-distorted image [14].

3.3.4 Deinterlacing artifacts

Since the start of television broadcast, video sequences have been transmitted in line interlaced format. Several different perceptual artifacts are related to this technology, ranging from the inversion of the top-field and the bottom-field, to the de-interlacing algorithms used in current display technologies. A technical overview has been provided in [16] while de-interlacing techniques have been subjectively compared in [72].

3.3.5 Spatial error concealment

Packet losses or bit inversions lead to missing content at the receiver side which is often replaced by previously transmitted content or by in-painting of surrounding regions. The result is often an isolated image region which does not fit to the surrounding perceptual information. Due to the prediction used in video coding, this artifact has a temporal duration as the image

region tends to grow as its blocks are used for further image prediction. This is particularly visible during a scene cut when mixture of different contents may occur.

3.3.6 Channel switches

Voluntary and involuntary channel switches form a special kind of artifacts because they may introduce content mixtures similar to error concealment artifacts or annoy the user by pauses in the transmission [34].

3.4 Classification of temporal degradations

3.4.1 Frame rate

The number of frames per second in video transmission is often less than the perceptual maximum for a given viewing angle. In [41] a model is proposed to estimate the normalized quality as a function of the frame rate. The model consists of an inverted exponential function and incorporates also a parameter that characterizes how fast the perceived quality reduces as the frame rate decreases, that depends on the content characteristics, such as the motion or the resolution. At the same spatial resolution, faster motion sequences have a higher falling rate. Also, for the same sequence, the QCIF resolution shows a higher falling rate than CIF.

3.4.2 Frame freezing and frame skipping

Video transmission over networks is often impaired by delays or outages. When video content cannot be decoded at a given time, the playback pauses and when it resumes, it may either continue with the next frame or skip a few frames to compensate for the previously paused delay time. This type of distortion may appear both in TCP and UDP transmission. In [50] the authors propose a metric to compute MOS as a function of the number of pauses, the average length of pauses that happened in the same temporal segment, a weighting factor which represents the degree of degradation that each segment adds to the total video degradation, the time period in seconds of each segment and the number of temporal segments of a video. It is interesting to note that they divide the video into segments and find the weighting factors for each segment solving an equations system. It was found that the initial video segment is more relevant, or it has more impairment weight in relation to other video temporal segments. Pauses at the beginning of the video have a higher negative effect on user QoE. In [71] the authors conclude that the perceptual impact of frame freeze and frame skip is highly content dependent and that viewers prefer a scenario in which a single but long freeze event occurs to a scenario in which frequent short freezes occur.

3.5 Classification of attention related indicators

3.5.1 Visual attention and saliency

Determining the most important image regions in a video content simplifies the spatial pooling but may also help in determining whether the observer is immersively attracted by a single object or whether he is freely scanning the complete image region. A typical example is an interview situation, when faces attract attention and degradations in surroundings are less noticeable. Predictions may therefore be improved by using saliency awareness [17, 38].

3.5.2 Visibility duration of objects

Newly appearing objects in a video scene require time for their cognitive understanding. During this time, the perception of artifacts is reduced [3, 17].

3.5.3 Forgiveness effect

An article about the temporal characterization of forgiveness effect in which they introduced a degree of blockiness in the videos [23] shows that the forgiveness effect is initially greater when good quality material follows a high level degradation compared to a low level degradation. However, with increasing periods of good quality material, subjects are more forgiving of low level degradation.

4 Existing video quality estimation algorithms

The previous section listed isolated perceptual artifacts and models that were proposed to estimate the severity of the artifact. Some of the models were implemented as algorithms and in some cases the corresponding references were provided for measuring the influence of an isolated artifact in a particular condition.

On the opposite end of the scale, there are algorithms which are meant to measure a certain group of artifacts that may occur in a realistic transmission scenario. These algorithms are a compromise between execution speed, model accuracy, and prediction accuracy. The latter two are distinguished by the appropriateness to model an isolated aspect of the human perception and the prediction performance of the model in the given scenario. For example, correctly modeling the effect of forgiveness may be important but as in most cases only 8–12 s of video were considered, the influence may be negligible for the model's performance.

Table 1 lists a selection of measurement methods from very simple models to complex models which underwent validation and standardization. As most of the validation has been performed within VQEG, the application areas or scopes were retained even outside of the standardization efforts as the same video sequences were used for performance evaluations. In VQEG-SD Phase I and Phase II, standard definition (SD) television was considered as scope without extreme low-bitrate coding conditions, as well as without frame rate reduction and without packet losses. In VQEG-MM Phase I, typical multimedia (MM) resolutions and conditions at that time were considered, i.e. Quarter Common Interchange Format (QCIF), CIF, and VGA. Very low bit-rate was included as well as frame rate reduction and network packet losses. In VQEG-HD Phase I High Definition (HD) content was degraded by similar conditions as used in VQEG-MM.

Most of the currently employed video quality measurement algorithms are FR or RR models as the performance of NR models may not be expected to be sufficient for industrial application.

For each of the models, the table shows to which extent a perceptual artifact has been taken into consideration. The scale follows the Absolute Category Rating scale. A value of five (5) indicates that the modeling is very elaborate, a value of four (4) indicates that substantial parts of the model are dedicated to measuring this particular artifact. In cases when the model takes the artifact into consideration but is not particularly targeted to its measure, a value of three (3) was given. Two (2) indicates that the model may take this degradation into consideration but there is no unique indicator included. One (1) indicates that the model handles the artifact while

Table 1 Video quality measurement algorithm and their indicators

Abbreviation		PSNR ATIS	FULLPARSE	NORM	TetraVQM	Standard A and B	PVQM	ATIS T1.TR.PP.76	ITU-T J.246 Annex A	ITU-T J.342
Reference		[29]	[49]	[43]	[4]	[63]	[24]	[57]	[27]	[28]
Algorithm type		FR	NR	NR	FR	FR	FR	FR	RR	RR
Scope limitations		Temporally aligned luminance components	Estimates MSE, not MOS. Assumes specific error concealment.	Estimates MSE, not MOS. Assumes specific error concealment.	VQEG-MM	Detection of contrast pattern	VQEG-SD Phase I	VQEG-SD Phase I	VQEG-MM	VQEG-HDTV
Signal processing	Image difference to reference	3	2	3	3	3	3	4	3	3
Perceptual: content related	CSF	-	-	-	-	5	2	2	-	-
	Spatial masking	-	-	-	-	-	-	4	-	2
	Temporal masking	-	-	-	4	-	-	-	-	-
Perceptual: spatial degradation	Blockiness	3	-	-	3	1	4	3	3	4
	Blurriness	3	-	-	3	1	3	3	3	3
	Geometric distortions	-	-	-	2	-	-	-	-	-
Perceptual: content replacement	Imperfect error concealment after transmission errors	-	-	-	3	-	-	-	2	2
Perceptual: temporal degradation	Frame rate	-	-	-	3	-	-	-	2	2
	Pausing	-	-	-	3	-	-	-	2	2
	Skipping	-	-	-	3	-	-	-	2	2
	Rebuffering	-	-	-	3	-	-	-	-	-
Perceptual: attention and cognitive aspects	Visual saliency	-	-	-	3	-	-	-	-	-
	Visibility duration of objects	-	-	-	4	-	-	-	-	3
	Recency	-	-	-	-	-		-	-	-
	Forgiveness	-	-	-	-	-		-	-	-
	Object recognition	-	-	-	2	-		-	-	-
Complex techn. artefacts	Interlacing	-	-	-	-	-	3	-	-	1
	Deinterlacing	-	-	-	-	-	-	-	-	-

estimating the quality related to a different artifact, the prediction performance may be limited. A dash (-) signifies that the artifact was not considered for the model according to its description.

In the case of bitstream-based objective video quality metrics, quality is estimated by analyzing the received video bitstream. As such, no full decoding is performed and different parameters are extracted from the (impaired) encoded video bitstream. In [2], parameters are extracted at the frame, macroblock and motion vector level to continuously estimate visual quality. These parameters are used to calculate a number of indicators such as error duration, error propagation, and frame freezing. Finally, the indicators are combined and temporal pooling is applied. Yang et al. [70] propose an NR bitstream-based metric, which measures video quality using three key factors: picture distortions resulting from quantization, quality degradation caused by network impairments, and temporal effects of the Human Visual System (HVS). Characteristics of the HVS such as spatial and temporal masking are also taken into account when calculating the picture quality. Both spatial and temporal pooling are applied in their quality assessment framework. Recently, ITU-T Study Group 12 (SG12) also published recommendation P.1202.1 describing parametric non-intrusive bitstream assessment of low resolution video streaming quality. High resolution video applications, including IPTV, are still under investigation.

In order to reliably measure spatial degradations such as blockiness and blurring, access to pixel data is required. In the case of hybrid quality metrics, both the encoded (impaired) video stream and the decoded video data are made available. This approach has, for example been used in [18] and [37] for measuring video quality by detecting perceptual artifacts. The authors in [30, 31, 48] extract different parameters from the received and decoded video stream to model the influence of packet loss on impairment visibility. Features are extracted to identify the spatial and temporal extent of the loss and combined with the mean square error of the initial error averaged over the macroblocks initially lost. The V-Factor [67] objective video quality metric inspects the Transport Stream (TS) and Packetized Elementary Stream (PES) headers, and encoded and decoded video signal to estimate perceived visual quality. V-Factor also considers content characteristics, compression mechanism, bandwidth constraints and network impairment such as jitter, delay and packet loss to measure video quality in real-time. Yamagishi et al. [69] extend a packet-layer model with content-based information extracted from the decoded video stream in order to estimate video quality per content type. As such, information is combined from the packet headers and video signals. Similarly, results in [32] also show that the overall prediction accuracy of a pure bitstream based objective video quality model can be significantly increased when pixel-based features are taken into account. From the bitstream, features are obtained at the slice, macroblock and motion vector level. These features are then combined with pixel-based features such as blockiness, blurriness, activity, predictability, and motion, edge, and color continuity.

VQEG's JEG is also particularly interested in the construction, validation and evaluation of hybrid bitstream-based objective video quality assessment.

5 Scope considerations for the combination of indicators

A critical question in the development of a video quality measurement algorithm concerns the choice of the indicators and their combination. While experts can usually perceptually distinguish between different artifact types as listed in Section 4, measurement algorithms are often not sufficiently discriminative. Each indicator algorithm usually has a certain artifact

for which it was designed and trained on. This artifact or class of artifacts may be said to be "in scope" for the algorithm. Other artifact types may be sufficiently close to be also predicted but with less accuracy, those shall be termed as being "in extended scope". The remaining artifacts should be considered as being "out of scope". For an example of this definition, the reader is referred to [5].

In an ideal measurement algorithm, each considered artifact has exactly one single indicator. Each such indicator performs perfectly "in scope", does not have an "in extended scope" range, and stays neutral, i.e. "no artifact present" when "out of scope" degradations occur.

In practice, these conditions do not hold true, and what is more important they have not been taken into consideration so far in model development. A comprehensive analysis of each indicator is often only available for "in scope" conditions, while the other conditions are not evaluated.

In some cases, it is algorithmically determined that an indicator will not measure a degradation, for example, an image based indicator will not measure temporal degradations. This is no longer true when the indicator is used as part of a complex algorithm because the algorithm may have side effects. In VQEG-MM for example, a PSNR variant [68] has been used on videos with temporal degradations such as pausing and skipping. The algorithm only estimated a constant time offset for the complete video sequence and thus non-matching frames were compared by PSNR in case of pauses or skips, enabling the algorithm to predict these degradations to some extent.

The determination of a particular scope for each indicator and for the overall model requires necessarily a validation process. This validation process usually uses data that was unknown to the model during its development. A possible alternative may be to evaluate each algorithm on the psychophysical databases that were used to develop the individual indicators and to learn about the behavior of the other "out of scope" or "in extended scope" indicators.

It should also be taken into consideration that some indicators may take binary precedence over others. Some may be useful only in a certain range of a wider application scope. For example, evaluating thresholds at contrast sensitivity may be used when the content is of very high quality while this indicator is turned off when a reduced frame rate has been detected.

Models often require intensive training and often the numerically optimized results of the training output are not comprehensible from a vision modeling point of view.

In Table 1 a list of algorithms was provided for which the design scope was fixed in testplan documents which were later used in standardization. The table itself lists the performance of the model with respect to the presence and the algorithmic description of each indicator. In order to determine the experimental scope of each indicator, extensive validation experiments would be required and the result of the individual indicators may be different from their combination due to the model's processing framework and the training performed.

One issue in this effort is the development of a common objective scale. Usually, each objective model is allowed a fitting to a subjective dataset before evaluation, in most cases linear, third order monotonic or sigmoidal fittings are employed. However, when the influence of indicators shall be combined and the rating scale shall extend over the range that may be measured in one single subjective experiment with sufficient precision, the development of an objective scale is required. For example, the Just Noticeable Difference scale [62] would allow for a subjectively and objectively meaningful absolute scale by specifying a percentage of detection probability of differences between images. This scale is compatible with data obtained from Paired Comparison experiments when using the Thurstone-Mosteller or Bradley-Terry model

[7, 58]. Their usage for video sequence would have to be analyzed. Several subjective experiments would be required to establish a link to other commonly used subjective scales and also for evaluating the combination of subjective experiments.

6 Towards a joint development of video quality measurement

In the video coding community, the most appropriate algorithms have long been identified and continuous work led to a succession of successful standards which often reduced the bit-rate by half at similar perceived quality. In video quality prediction, various standards and algorithms exist that have been developed in an isolated manner. Evolution is difficult to measure.

From the previous sections the following process for a joint development of video quality estimation may be derived:

1 Creation of psychophysical stimuli databases which allow the measurement of isolated vision and/or quality aspects. Results both in terms of video sequences of the stimuli as well as subjective assessment results need to be made available.
2 Creation of natural and synthetic video databases impaired by isolated perceptual artifacts also made publicly available.
3 Development of a common objective quality scale and tools to realign currently employed subjective quality assessment scales to this common scale.
4 Test, development and refinement of individual indicators, identifying each time their performance "in scope" with the previously developed databases, their performance "in extended scope" and their neutrality in "out of scope" conditions.
5 Development of frameworks of combination of several indicators with respect to their identified scope performance.

This outline covers only the most important work packages. For example, the correct temporal, spatial and color alignment of video sequences is not mentioned as well as the complex influence from object recognition and visual saliency. The performance also depends on the temporal distribution of the artifacts, and content features such as spatial complexity, temporal complexity, camera and object motion, and scene changes. A major problem of the development of an algorithm is also the interdependency of the chosen indicators, some may be meant to measure second or third order effects which are partially already covered by "extended scopes" of others.

Bridging the gap between perception modeling and computer algorithms requires international cooperation and structured research.

Recently, the VQEG JEG Hybrid group has started work towards the development of a Hybrid NR video quality measurement that aims at establishing a collaborative way of developing objective video quality measurement. Their work will take into consideration the above-mentioned five steps. As previous work on various steps is available and iterations are required the work will advance in parallel. VQEG's JEG is free and open to everyone, both from academia as well as private industries. No subscription fees are involved for joining VQEG JEG. Contributions can be made concerning every step involved in subjective and objective video quality assessment. VQEG JEG also collaborates with other entities such as the COST IC1003 "Qualinet" network.

VQEG JEG has chosen to develop a Hybrid algorithm. A toolchain is available that allows to parse the bitstream into an XML based file format that is human readable. All kinds of measurement algorithms and scopes are considered. The current main focus is on Hybrid NR model indicators.

As compared to FR models, hybrid NR allows the measurement at the client side and retrieval of bitstream information in conjunction with the analysis of the decoded video sequence may result in higher prediction accuracy. As an example, bitstream information may help localizing packet delays or packet losses in transmission while the decoded video may be used to estimate the efficiency of error concealment. Theoretical limits such as homogeneous and inhomogeneous transcoding will be considered. For example, a video sequence encoded at a low bitrate first and then transcoded to a higher bit-rate, thus encoding mostly the previous coding errors, is difficult to detect in a pure bitstream model. Industrial application difficulties such as encryption may be considered and simulated. While NR models may have difficulties in estimating quality of particular non-natural contents, notably cartoons or drawings, their quality may be estimated by the bitstream analysis. This interrelationship between bitstream analysis and NR estimation immediately leads to the above-mentioned integration of indicators with respect to their individual scope.

7 Conclusions

The research on subjective and objective video quality assessment is often considered an interdisciplinary task. A lot of effort has been spent by different research communities to identify and measure perceptual attributes of the HVS, to develop, train, verify, and validate individual indicators, or to develop perceptually meaningful spatial and temporal pooling algorithms. Data mining and machine learning has been applied to combine the different predictions into a global quality score. It has however been shown that some of the isolated indicators may not be suitable for the complete application range or they may behave erroneously when confronted with degradations outside of their scope. In this context, the importance of creating various subjectively evaluated datasets has been emphasized. Each such databases may contain either psychophysical stimuli or an isolated degradation type or an overall scope evaluation. The combination of the subjective voting given on each individual database needs to be tackled leading to the notion of an objective video quality scale. The individual indicators may then be analyzed for the degradations which they are supposed to measure, i.e. “in scope” degradations, which they may measure with less prediction performance “extended scope”, and their neutral behavior in all other conditions “out of scope”. VQEG JEG’s Hybrid group is working towards organizing collaboration towards these goals.

Acknowledgments The authors would like to thank the members of the Video Quality Experts Group (VQEG), in particular Arthur Webster, for all the support they have received. Furthermore, the work of Mikołaj Leszczuk is supported by The Polish National Centre for Research and Development (NCRD) under Grant No. SP/I/1/77065/10 by the Strategic Scientific Research and Experimental Development programme: “Interdisciplinary System for Interactive Scientific and Scientific-Technical Information”. The work of Kjell Brunnström has been supported by the Swedish Governmental Agency for Innovation Systems (Vinnova), which is hereby gratefully acknowledged.

References

1. Ahumada AJ Jr, Beard BLJ, Eriksson R (1998) Spatio-temporal image discrimination model predicts temporal masking functions. In: Rogowitz B, Pappas TN (eds) Proc. SPIE Human Vision and Electronic Imaging III: 3299. SPIE, Bellingham, pp 152–166

2. Argyropoulos S, Raake A, Garcia M-N, List P (2011) No-reference video quality assessment for SD and HD H.264/AVC sequences based on continuous estimates of packet loss visibility. Third International Workshop on Quality of Multimedia Experience (QoMEX)
3. Barkowsky M, Eskofier B, Bialkowski J, Kaup A (2006) Influence of the presentation time on subjective votings of coded still images. ICIP, 429–432
4. Barkowsky M, Bialkowski J, Eskofier B, Bitto R, Kaup A (2009) Temporal trajectory aware video quality measure. IEEE J Sel Top Signal Proc 3(2):266–279
5. Barkowsky M, Staelens N, Janowski L, Koudota Y, Leszczuk M, Urvoy M et al (2012) Subjective experiment dataset for joint development of hybrid video quality measurement algorithms. In: QoEMCS 2012—Third Workshop on Quality of Experience for Multimedia Content Sharing (pp. 1–4). Berlin, Allemagne
6. Barten P (1999) Contrast sensitivity of the human eye and its effects on image quality. SPIE, Bellingham
7. Bradley RA (1984) 14 paired comparisons: some basic procedures and examples. Handb Stat 4: 299–326
8. Bruce V, Georgeson MA, Green PR, Georgeson MA (2003) Visual perception: physiology, psychology and ecology. Psychology Press, Hove
9. Brunnström K, Schenkman BN (2002) Comparison of the predictions of a spatio-temporal model with the detection of distortion in small moving images. Opt Eng 41(3):711–722
10. Brunnström K, Eriksson R, Ahumada AJ Jr (1999) Spatio-temporal discrimination model predicting IR target detection. In: Rogowitz B, Pappas TN (eds) Proc. SPIE Human Vision and Electronic Imaging IV: 3644. SPIE, Bellingham, pp 403–410
11. Brunnström K, Wang K, Sedano I, Barkowsky M, Kihl M, Aurelius A, Le Callet P, Sjöström M (2012) 2D no-reference video quality model development and 3D video transmission quality. Proc. 6th Inter. Workshop on Video Processing and Quality Metrics for Consumer Electronics: Scottsdale, AZ, USA
12. Burbeck CA, Kelly DH (1980) Spatiotemporal characteristics of visual mechanisms: excitatory-inhibitory model. J Opt Soc Am 70(9):1121–1126
13. Chikkerur S, Sundaram V, Reisslein M, Karam LJ (2011) Objective video quality assessment methods: a classification, review, and performance comparison. IEEE Trans Broadcast 57(2):165–182
14. D'Angelo A, Zhaoping L, Barni M (2010) A full-reference quality metric for geometrically distorted images. IEEE Trans Image Proc 19(4):867–881
15. Daly SJ (1998) Engineering observations from spatiovelocity and spatiotemporal visual models. 180–191
16. De Haan G, Bellers EB (1998) De-interlacing: an overview. Proc IEEE 86(9):1839–1857
17. Engelke U, Barkowsky M, Le Callet P, Zepernick H-J (2010) Modelling saliency awareness for objective video quality assessment. In: IEEE international workshop on Quality of Multimedia Experience QoMEX. Trondheim, Norway
18. Farias M, Carvalho M, Kussaba H, Noronha B (2011) A Hybrid Metric for Digital Video Quality Assessment. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp 1–6
19. Foley JM (1994) Human luminance pattern-vision mechanisms: masking experiments require a new model. J Opt Soc Am A 11(1710):1719
20. Fredericksen RE, Hess RF (1999) Temporal detection in human vision: dependence on spatial frequency. J Opt Soc Am A 16(11):2601–2611
21. Girod B (1989) The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals. Visual Processing, and Digital Display, SPIE, 178–189
22. Goldstein BE (2006) Sensation and perception, 7th edn. Wadsworth Publishing
23. Hands DS (2001) Temporal characterisation of forgiveness effect. Electron Lett 37(12):752–754
24. Hekstra AP, Beerends JG, Ledermann D, de Caluwe FE, Kohler S, Koenen RH et al (2002) PVQM—a perceptual video quality measure. Elsevier, Signal Processing: Image Communications 17:781–798
25. International Telecommunication Union, Telecommunication standardization sector ITU-T (2012) Parametric non-intrusive assessment of audiovisual media streaming quality. ITU-T Rec P.1201
26. International Telecommunication Union, Telecommunication standardization sector ITU-T (2012) Parametric non-intrusive bitstream assessment of video media streaming quality. ITU-T Rec P.1202
27. International Telecommunications Union (ITU) (2008) Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference. ITU Recommendation J.246
28. International Telecommunications Union (ITU) (2011) Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal. ITU Recommendation J.342
29. NTIA / ITS (2001) A3: Objective Video Quality Measurement Using a Peak-Signal-to-Noise-Ratio (PSNR) Full Reference Technique. ATIS T1.TR.PP.74-2001
30. Kanumuri S, Cosman P, Reibman A, Vashampayan V (2006) Modeling packet-loss visibility in MPEG-2 video. IEEE Trans Multimedia 8(2):341–355

31. Kanumuri S, Subramanian S, Cosman P, Reibman R (2006) Predicting H.264 packet loss visibility using a generalized linear model. In: International Conference on Image Processing, pp 2245–2248
32. Keimel C, Habigt J, Diepold K (2012) Hybrid no-reference video quality metric based on multiway PLSR. In: Proceedings of the 20th European Signal Processing Conference pp 1244–1248
33. Koenderink JJ, van Doorn AJ (1979) Spatiotemporal contrast detection threshold surface is bimodal. Opt Lett 4(1):32–34
34. Kooij R, Nicolai F, Ahmed K, Brunnström K (2009) Model validation of channel zapping quality. Proc SPIE 7240:72401
35. Le Callet P, Möller S, Perkis A (eds) (2012) Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.1
36. Legge GE, Foley JM (1980) Contrast masking in human vision. Opt Soc Am 70(12):1458–1471
37. Liao N, Chen Z (2011) A packet-layer video quality assessment model with spatiotemporal complexity estimation. EURASIP J Image Video Proc 2011(1):1–13
38. Liu T, Feng X, Reibman A, Wang Y (2009) Saliency inspired modeling of packet-loss visibility in decoded videos. VPQM
39. Lubin J (1993) The use of psychophysical data and models in the analysis of display system performance. In: Watson AB (ed) Digital images and human vision. Cambridge, MIT Press, pp 163–178
40. Lubin J (1995) A visual discrimination model for imaging system design and evaluation. In: Vision models for target detection and recognition. World Scientific Publishing Company, Incorporated, River Edge, pp 245–283
41. Ma Z, Xu M, Ou Y-F et al (2012) Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications. IEEE Trans Circ Syst Video Technol 22(5):671–682
42. Malacara D (2001) Color vision and colorimetry: theory and applications (vol. PM105). SPIE Press
43. Naccari M, Tagliasacchi M, Tubaro S (2009) No-reference video quality monitoring for H.264/AVC coded video. IEEE Trans Multimedia 11(5):932–946
44. Peli E (1990) Contrast in complex images. J Opt Soc Am A 7(10):2032–2040
45. Poynton C (1998) Frequently Asked Questions about Gamma. Retrieved April 22, 2013, from http://www.poynton.com/PDFs/GammaFAQ.pdf
46. Poynton C (2003) Digital video and HDTV algorithms and interfaces. Morgan Kaufmann Publisher, Elsevier Science, San Francisco, ISBN 1-55860-792-7
47. Poynton C (2006) Color FAQ Frequently Asked Questions Color. Retrieved April 22, 2013, from http://www.poynton.com/PDFs/ColorFAQ.pdf
48. Reibman A, Kanumuri S, Vaishampayan V, Cosman P (2004) Visibility of individual packet losses in MPEG-2 Video. In: International Conference on Image Processing, pp 171–174
49. Reibman AR, Vaishampayan VA, Sermadevi Y (2004) Quality monitoring of video over a packet network. IEEE Trans Multimed 6(2):327–334
50. Rodriguez DZ, Abrahao J, Begazo DC et al (2012) Quality metric to assess video streaming service over TCP considering temporal location of pauses. IEEE Trans Consum Electron 58(3):985–992
51. Romaniak P, Janowski L, Leszczuk M, Papir Z (2012) Perceptual quality assessment for H.264/AVC compression. In: Consumer Communications and Networking Conference (CCNC), 2012 IEEE, pp 597–602
52. Staelens N, Sedano I, Barkowsky M, Janowski L, Brunnström K, Le Callet P (2011) Standardized toolchain and model development for video quality assessment—the mission of the Joint Effort Group in VQEG. In: Proceedings of 2011 Third International Workshop on Quality of Multimedia Experience (QoMEX). Mechelen, Belgique
53. Strohmeier D, Jumisko-Pyykkö S, Kunze K (2010) New, Lively, and Exciting or Just Artificial, Straining, and Distracting—a Sensory Profiling Approach to Understand Mobile 3D Audiovisual Quality. Fifth International Workshop on Video Processing and Quality Metrics (VPQM)
54. VQEG Tools and Subjective Labs Setup (2010) Project Homepage of the VQEG STL, [Online]. Available: http://vqegstl.ugent.be
55. Takahashi A, Hands D, Barriac V (2008) Standardization activities in the ITU for a QoE assessment of IPTV. IEEE Commun Mag 46(2):78, 84
56. Tam WJ, Stelmach LB, Wang L, Lauzon D, Gray P (1995) Visual masking at video scene cuts. Human Vision, Visual Processing, and Digital Display VI, SPIE, 111–119
57. Tektronix Inc. (2001) A4: Objective Perceptual Video Quality Measurement Using a JND-Based Full Reference Technique. ATIS T1.TR.PP.76-2001

58. Thurstone LL (1927) A law of comparative judgment. Psychol Rev 34(4):273
59. van den Branden Lambrecht CJ (1996) Perceptual Quality Measure using a Spatio-Temoral Model of the Human Visual System. Proc SPIE Digital Video Compression: Algorithms and Technologies, San Jose, CA, USA, Vol. 2668, (pp 450–461)
60. Video Quality Experts Group (2007) Project Homepage of VQEG Joint Effort Group—Hybrid. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/jeg/jeg.aspx
61. Video Quality Experts Group (VQEG) (2003) Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II. VQEG Final Report of FR-TV Phase II Validation Test, Video Quality Experts Group (VQEG)
62. Watson AB (2002) Draft Standard for the Subjective Measurement of Visual Impairments in Digital Video Using a Just Noticeable Difference Scale. P1486, Revision D06
63. Watson AB, Ahumada AJ Jr (2006) A standard model for foveal detection of spatial contrast. J Vis 5(9):717–740
64. Watson B, Solomon JA (1997) A model of visual contrast gain control and pattern masking. J Opt Soc Am A 14:2379–2391
65. Westen SJP, Lagendijk RL, Biemond J (1996) Adaptive spatial noise shaping for DCT based image compression. ICASSP 4:2124–2127
66. Winkler S (2000) Vision models and quality metrics for image processing applications. PhD Thesis no 2313, Ecole Polytechnique Fédérale de Lausanne
67. Winkler S, Mohandas P (2008) The evolution of video quality measurement: from PSNR to hybrid metrics. IEEE Trans Broadcast 54(3):660–668
68. Wolf S, Pinson MH (2009) Reference Algorithm for Computing Peak Signal to Noise Ratio (PSNR) of a Video Sequence with a Constant Delay. ITU-T Contribution COM9-C6-E
69. Yamagishi K, Kawano T, Hayashi T (2009) Hybrid video-quality-estimation model for IPTV services. Global Telecommunications Conference (GLOBECOM)
70. Yang F, Wan S, Xie Q, Wu HR (2010) No-reference quality assessment for networked video via primary analysis of bit stream. IEEE Trans Circ Syst Video Technol 20(11):1544–1554
71. Yining Q, Mingyuan D (2006) The effect of frame freezing and frame skipping on video quality. In: Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Pasadena
72. Zhao M, de Haan G (2005) Subjective evaluation of de-interlacing techniques. Proc SPIE 5685:683–691

Marcus Barkowsky received the Dr.-Ing. degree from the University of Erlangen-Nuremberg in 2009. Starting from a deep knowledge of video coding algorithms his Ph.D. thesis focused on a reliable video quality measure for low bitrate scenarios. Special emphasis on mobile transmission led to the introduction of a visual quality measurement framework for combined spatio-temporal processing with special emphasis on the influence of transmission errors. He joined the Image and Video Communications Group IRCCyN/IVC at the University of Nantes in 2008, and was promoted to associate professor in 2010. His activities range from modeling effects of the human visual system, in particular the influence of coding, transmission, and display artifacts in 2D and 3D to measuring and quantifying visual discomfort and visual fatigue on 3D displays using psychometric and medical measurements. He currently co-chairs the VQEG “3DTV” and “Joint Effort Group Hybrid” activities.

Iñigo Sedano is Telecommunications Engineer by the University of Deusto (Bilbao, 2005) and Master in Information Technologies and Communications in Mobile Networks by the University of the Basque Country (Bilbao, 2007). In the years 2005–2006 he worked for DeustoTech (Deusto Institute of Technology) at the University of Deusto in the field of ambient intelligence and telematic solutions. Since October 2006 he works for the Division ICT-European Software Institute of Tecnalia and has participated in European research projects related to broadband networks (PlaNetS, Banits2, TRAMMS). Between November 2009 and May 2011 he completed a research fellowship at Acreo Research Institute, Sweden about methods to evaluate the video quality. From November 2011 he participates in various Spanish research projects in the area of scalable video coding at Tecnalia Research & Innovation.

Kjell Brunnström, Ph.D. is a Senior Scientist at Acreo Swedish ICT AB and Adjunct Professor at Mid Sweden University. He is an expert in image processing, computer vision, image and video quality assessment having worked in the area for more than 25 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 having reviewed a number of scientific articles for international peer-reviewed 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 Ph.D. and M.Sc students. Currently, he is leading standardisation activities for video quality measurements as Co-chair of the Video Quality Experts Group (VQEG). He is one of two Swedish MC representative in EU-Cost action QUALINET. His current research interests are in Quality of Experience for visual media in particular video quality assessment both for 2D and 3D, as well as display quality related to the TCO requirements.

Mikołaj Leszczuk, PhD is an assistant professor at the Department of Telecommunications, AGH University of Science and Technology ((AGH-UST), Krakow, Poland). He received his M.Sc. in Electronics and Telecommunications in 2000 and PhD degree in Telecommunications in 2006, both from AGH-UST. He is currently teaching Digital Video Libraries, Information Technology and Basics of Telecommunications. In 2000 he visited Universidad Carlos III de Madrid (Madrid, Spain) for a scientific scholarship. During 1997–1999 he was employed by several Comarch holding companies as a Manager of Research and Development Department, President of the Management and Manager of the Multimedia Technologies Department. He has participated actively as a steering committee member or researcher in several national and European projects, including: INDECT, BRONCHOVID, GAMA, e-Health ERA, PRO-ACCESS, Krakow Centre of Telemedicine, CONTENT, E-NEXT, OASIS Archive, and BTI. He is a member of the VQEG (Video Quality Experts Group) Board and a co-chair of VQEG QART (Quality Assessment for Recognition Tasks) Group. His current activities are focused on e-Health, multimedia for security purposes, P2P, image/video processing (for general purposes as well as for medicine), the development of digital video libraries, particularly video summarization, indexing, compression and streaming subsystems. He has been a chairman of several conference sessions. He is a member of IEEE Society since 2000. He has served as an expert for the European Framework Programme and the Polish State Foresight Programme as well as a reviewer for several conferences publications and journals.

Nicolas Staelens obtained his Master's degree in Computer Science at Ghent University (Belgium, 2004). In 2006, he joined the Internet Based Communication Networks and Services (IBCN) group at Ghent University where he received a Ph.D. degree in Computer Science Engineering in February 2013. The topic of his dissertation was "Objective and Subjective Quality Assessment of Video Distributed over IP-based Networks". Currently, he is still working at the same university where his main research focuses on assessing the influence of network impairments on perceived audiovisual quality.

As of 2007, he is also actively participating within the Video Quality Experts Group (VQEG) and is currently co-chair of theTools and Subjective Labs support group and the JEG-Hybridproject.