

# REAL-TIME VIDEO QUALITY MONITORING FOR MOBILE DEVICES

*Tao Liu, Glenn Cash, Wen Chen, Chunhua Chen, Jeffrey Bloom*

Dialogic Research, Inc., Eatontown, NJ  
{first.last@dialogic.com}

## ABSTRACT

Due to the explosion of various video applications and services it is becoming increasingly important to accurately measure video quality in real-time. Since reference information is absent in many practical situations, video quality assessment with No-Reference (NR) measures is attracting increasing attention from both industry and academia. However, there are very few commercial quality assessment systems for mobile devices available in today's market.

After describing a set of applications, we briefly survey the state-of-the-art in NR video quality assessment. Finally, we present a real-time quality monitoring system for video transmitted to mobile devices and describe a proof-of-concept implementation. In this system, the quality of video at different nodes and terminals in a transmission channel can be monitored and tracked through a remote software interface. These video quality measurements can also provide valuable information for network diagnosis and quality-scalable service planning.

**Index Terms**— video quality assessment, H.264/AVC, No-reference quality measurement

## 1. INTRODUCTION

As the volume of video content processed and transmitted over communications networks increases, there is a corresponding increase in the need for accurate and real-time video quality assessment tools. These tools will be relevant for potential video applications and services, including applications involving video delivered to mobile devices. Some of these applications include end-user quality assurance, in-network quality monitoring of service level agreement compliance, and management of video quality for tiered services. However, although the need for video quality assessment tools is growing, there are not many commercial systems available.

The most commonly used methods for assessing visual quality are designed to predict subjective quality ratings on a set of training data. Many of these tools rely on access to an original undistorted version of the video under test. There has been significant progress in the development of such tools. However, they are not directly useful for many of the

new video applications and services in which the quality of a target work must be assessed without access to a reference. For these cases, no-reference (NR) models are more appropriate. Development of NR visual quality is a challenging research problem partially due to the fact that the artifacts introduced by different transmission components can have dramatically different visual impacts and the perceived quality can largely depend on the underlying video content.

In this paper, we address both of the above issues. The issue of transmission component artifact variety is approached by considering the change in measured quality across the network. To address the issue of content dependence, the measurement includes an analysis of the underlying content.

This paper begins with a discussion of some applications of quality assessment in the context of video delivery. This is followed by a brief survey of some NR video quality measures that are suitable for these applications. Finally, we describe a real-time quality monitoring demonstration system, designed and implemented as a proof-of-concept for the use of video quality measurement in real-time video delivery to smartphones.

## 2. QUALITY ASSESSMENT IN VIDEO SERVICES AND APPLICATIONS

The variety of video applications and services has been steadily growing. They include more mature services such as broadcast television, pay-per-view, and video on demand, as well as newer models for delivery of video over the internet to computers and over telephone systems to mobile devices such as smart phones. Niche markets for very high quality video for telepresence are emerging as are more moderate quality channels for video conferencing. As these services and applications become more widely deployed, new business models are emerging and the need for accurate, and in many cases real-time, video quality assessment is becoming increasingly important.

Perhaps the most obvious use of automated quality assessment is in quality assurance of end-user video. Today, network operators and content providers do not know when their users are receiving video of unacceptably low quality. The degradation in quality could be due to network errors or poor compression (perhaps a faulty

transcoder or one compressing at too low a bitrate). With visibility into the end-user quality, problems can be identified and possibly corrected. At the least, the customer relationship can be better managed with objective confirmation of the delivery of video at insufficient quality levels.

A second important use of automated quality assessment is as an objective measurement for service level agreements (SLA). Video is not passed directly from the content owner to the mobile device, but it is handled by a number of different parties before it reaches the viewer, such as access portals, content delivery networks (CDN's), backbone services, and one or more cellular operators. Each party may modify the video stream (typically a transcoding or transrating) for efficient transport through their specific network and hence potentially degrade the quality of original source video.

At each handoff point, the companies involved have agreements describing the level of service they will provide. This is often in terms of network availability, in addition to other Quality of Service (QoS) measures such as error rates, latency, and packet delivery jitter. These measures count all packet losses equally and do not capture perceptual quality degradations due to transcoding. Thus, the consumer can receive poor quality video even when all SLAs are met. SLAs for video content that include automated quality assessment measures will improve the perceptual quality of the video received by the consumer.

Some internet service providers and mobile service providers offer multiple tiers of service. Consumers can receive a higher tier of service by paying more. Often, these tiers are based on bandwidth (more money buys a larger maximum bitrate or maximum number of bytes per month). Consumers, however, don't think in terms of bytes per month. An alternative method would be to base the tiers on the quality of video delivered. A base tier would deliver a basic, low resolution, highly compressed video. At higher tiers, the video quality would improve. This is analogous to providing HD video at a higher rate than SD. In order to control the quality of the delivered video (just high enough to meet the expectations of the tier), good automated quality assessment measures are needed.

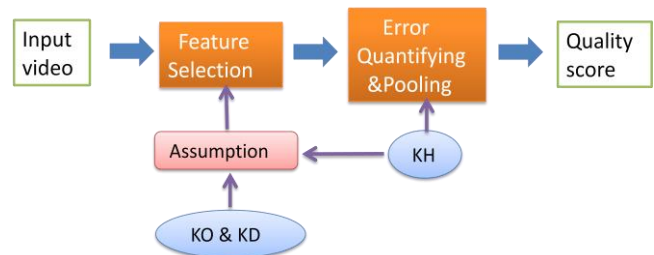
Networks can be designed to provide target QoS levels based on packet loss rates, latency, and packet delivery jitter and some networks are instrumented to enable monitoring of these values. But these measurements provide a very inaccurate estimate of the perceptual video quality. Incorporation of automated measures of perceptual quality as a supplement to traditional QoS measures can greatly improve the performance of networks that transport video content.

In many of the applications described above, the perceptual quality of the video must be estimated without reference to any other version of the video content. Objective, automated assessment tools that meet this

requirement are referred to as *no-reference* (NR) video quality measures.

### 3. STATE OF THE ART: NO-REFERENCE VIDEO QUALITY MEASURES

In the field of video quality measurement, techniques that rely on access to a reference are much more mature and accurate than no-reference models. One common approach to developing no-reference models is to make some assumptions about the reference and use these as part of the model. Generally speaking, there are two steps when designing no-reference measures: identify the video distortion and quantify the distortion. In the first step, artifacts in the processed video are estimated and differentiated from the features of the original, distortion-free version. This identifying process is usually dependent on assumptions about features or distortion characteristics. Therefore, some prior knowledge about what the original distortion-free video should look like (KO), knowledge about what kinds of distortion are under consideration (KD), and knowledge about the impact of the properties of the human visual system (HVS) on perceptual video quality (KH) are necessary[2]. In the step of quantifying the identified distortions, KH is also used to link the distortions with the associated subjective impact on video quality. The diagram in Figure 1 summarizes these two processes.



**Figure 1. No-reference quality measurement design**

One of the major challenges in designing accurate quality measures is the large number of artifacts that could possibly occur during video transmission. This makes it a difficult task to devise a generic quality measure. This is especially true for no-reference quality measures because potential artifacts and their characteristics can only be estimated, not measured. Therefore, a “divide-and-conquer” approach is often adopted. Different models are designed to detect and measure specific artifacts, or impairments. A number of these models are then combined so that many impairments can be considered simultaneously.

Among various forms of artifacts, the most commonly studied, and those most relevant for the delivery of mobile video, are spatial coding artifacts, (e.g. blurriness and blockiness), temporal induced artifacts, and packet-loss related artifacts. In the remainder of this section, we will

discuss no-reference measures for these four types of artifacts.

### 3.1. Spatial Distortion: Blurriness

Blurriness introduced during delivery of video is usually caused by loss of high frequency components during compression. The strong deblocking filter in the H.264 decoder and some other image post-processing such as noise reduction, can also generate blurriness. Since blurriness is highly content-dependent and without fixed patterns, the distinction between image features and artifacts is often ambiguous. Hence, blurriness is difficult to assess.

Based on the idea that blurriness usually smoothes the sharp transition across image edges, Marziliano et al. [3] proposed a blurriness measure based on the width of edges. If a pixel is on an edge, the width is based on the distance between a local maximum on one side of the edge and a local minimum on the other side. Ferzli and Karam [4] extended this work by incorporating contrast masking into the edge width calculation. One drawback of the Marziliano et al. algorithm, as recognized in [3] is the reliability of edge detection in the presence of blurriness. For this reason, Liu et al. [5] applied a similar idea, but transition sharpness is measured only at macro-block boundaries. This method was only validated with H.264 video.

Blurriness can also be measured by characterizing the variance of the statistics impacted by blurriness. Some statistics considered include image variance, gradient, Laplacian, autocorrelation, kurtosis, frequency, and entropy histograms [4]. This approach is simple and does not require edge detection, but the resulting measures have limited accuracy over different video contents.

### 3.2. Spatial Distortion: Blockiness

Blockiness is introduced during video communications by lossy, block-based compression. Because each block is quantized independently, a discontinuity can be introduced at the block boundaries.

Two observations that guide many blockiness detection models are (1) blockiness will usually cause image discontinuities in both horizontal and vertical directions, and (2) the occurrences of blockiness follows a grid pattern. Methods have been proposed in both the spatial and frequency domains. One advantage of methods that work in the block-DCT domain is that they often allow assessment of blockiness in a compressed bitstream without fully decoding the video.

In the spatial domain, Wang et al. [6] quantified discontinuities by averaging the intensity differences across block vertical and horizontal boundaries. Babu and Perkiş [7] applied activity masking during the calculation of block-edge gradients which were used as evidence of perceptual discontinuity. In the frequency domain, Wang et al. [8] treated distorted images as the proposition of original and

periodic step noise signals. From this, they detected peaks in the power spectrum of realigned 1-D image signals.

A common problem associated with detection of blockiness is that while its occurrence may follow the pattern of a grid, the original position of this pattern could be shifted because of possible transcoding or resizing that occurred after the blocking artifacts were introduced. To address this problem, a grid localization process was proposed by Mujis and Kirenko [9] where differences across neighboring lines were accumulated and periodic peaks of these differences were detected as locations of a blockiness grid.

### 3.3. Temporal Distortions

Unlike the studies on spatial artifacts, where numerous tools and results obtained for image quality assessment can be extended to video, there has been relatively little work focusing on the impact of temporal artifacts on video quality. However, temporal artifacts play an equally important role in affecting perceived quality of video.

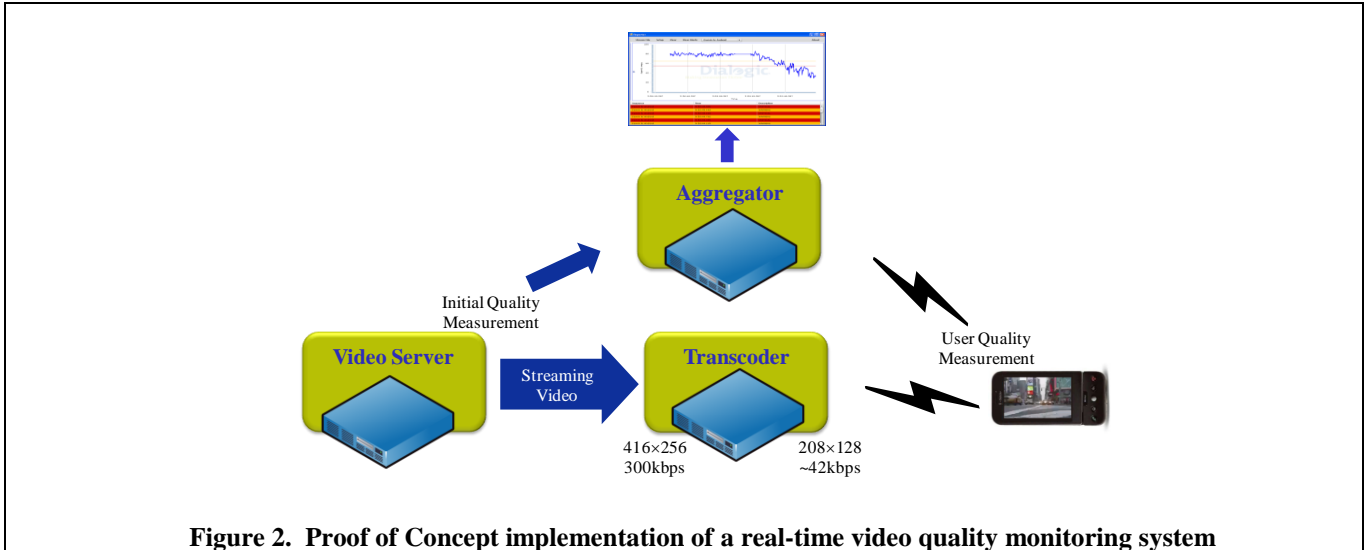
Of the studies that do exist, they are essentially limited to investigations of the relationship between video quality and frame rate. For this, there is a rich history in television and film where 24, 25, or 30 frames per second have been found to be sufficient to provide a high quality viewing experience.

In a recent review of the effect of frame rate on human perception, Chen and Thropp found that a minimum of 15 frames per second is needed to achieve a minimum level of viewer satisfaction [10]. They found that the exact acceptable frame rate is dependent on the video content and the underlying application and that it also varies greatly from one viewer to the next. The work by Yadavalli et al. also investigated the preferred frame rate for different types of video content [11].

Works by Lu et al. [12] and Yang et al. [13] proposed quality measures that include consideration of the effect of frame rate. They used different mapping functions to predict the effect of frame rate on perceptual video quality. Although these two models were shown to correlate well with subjective ratings for their target video scenarios, their models were solely dependent on frame rate without explicitly considering the impact of other temporal features, like motion, on perceptual video quality. In response, Ou et al. [14] characterized the relationship between perceptual video quality and frame rate with a logistic function whose parameters are modulated by video motion activity.

### 3.4. Packet Loss

Depending on the compression standard and the resolution of the imagery, a lost packet usually represents the loss of a slice of data (a number of horizontally adjacent blocks, sometimes affecting multiple adjacent rows of blocks) or the loss of an entire frame of data. In both cases, the decoder must attempt to approximate the lost data with certain error



**Figure 2. Proof of Concept implementation of a real-time video quality monitoring system**

concealment methods, and hence usually spatial artifacts can be introduced. Compression schemes that rely on motion compensation may then use the approximated data as a reference for prediction of data in subsequent frames. Thus, propagation artifacts can be introduced in frames that follow the lost packet. In general, errors can continue to propagate in time until the state is reset with a frame of data that does not depend on previous frames (e.g. an I-frame).

The visual artifacts introduced by packet loss can be dramatically different depending on a combination of a variety of factors including the video content itself (texture, brightness, amount of motion, etc.), the compression standard used, and the transmission protocol. For these reasons, summarized by the statement that all packets are not equally important to video quality, a simple measurement of packet loss rate does not accurately predict perceptual video quality. More sophisticated, perceptually-based measurements are needed.

Packet losses can cause both spatial and temporal distortions and clues to the existence of these artifacts can be found in discontinuities between the approximated slice or frame and its spatial and temporal neighbors. Some work in this area concentrates on detecting the loss of a slice or frame by looking at the pixel data while other work concentrates on assessing the impairment introduced by a known loss.

Pastrana-Vidal and Gicquel detected the loss of a frame by investigating motion fluidity. The quality of packet loss impaired video is then approximated with a psychovisual temporal pooling function [15]. Wolf detected the loss of frame data which has been concealed by repeating the last frame. This was done by examining the motion energy time history with an adaptive threshold [16]. Spatially, Babu et al. counted the number of pixels along the edges where gradient differences are high [17]. Slice Boundary Mismatch (SBM), proposed by Rui et al., considered the strength of continuities, together with its length, as an

indication of packet loss [18]. Reibman and Poole estimated the quality of packet loss impaired video from the perspective of the visibility of packet loss. A generalized linear model was used to leverage a bundle of quality metrics that can be easily extracted or estimated directly from the bitstream [19]. Finally, Liu assessed perceptual quality of videos affected by packet losses by incorporating several HVS properties, e.g. visual saliency, masking effects, and forgiveness effect [20].

In addition to the models developed for specific distortions, there are investigations into generic quality measurement which can predict the quality of video affected by multiple distortions. In 2007, ITU-T standardized such a no-reference quality measure. In ITU-T Recommendation G.1070 video quality is predicted from a few parameters in the bitstream [21]. The goal of current efforts within the ITU and the Video Quality Experts Group (VQEG) is to extend this work to combine parameters from the bitstream with measurements from the decoded image data [22]. This work is being developed specifically for multimedia streaming.

#### **4. REAL-TIME VIDEO QUALITY MONITORING SYSTEM FOR MOBILE DEVICES**

In order to illustrate some of the ideas described in Section 0, we have implemented a real-time video quality monitoring system, shown in Figure 2. This is not intended to be a deployable system; rather it is intended to demonstrate the technical feasibility of using NR video quality measures in real-time video delivery to a mobile device.

The video content is stored on a video server in approximately CIF resolution<sup>1</sup>, compressed with H.264/AVC at 300 kb/s. This resolution and bitrate is typical of Internet video. The content is then transcoded to a resolution and bitrate more typical of mobile video: approximately QCIF resolution at about 42 kb/s, again with H.264/AVC. Finally, the content is streamed to an HTC Android Dev Phone.

A video quality measurement (VQM) software agent has been installed on the phone. As the video frames are decoded, they are analyzed by the agent and the perceptual quality of the video is estimated. This agent currently uses some of the state-of-the-art NR techniques described in Section 3 and can, in general, be as sophisticated as the application will allow. For example, the VQM agent can analyze the bitstream as input to the decoder and/or the pixels as output from the decoder. If integration into the decoder is possible, the VQM can also consider coded information such as frame rate, frame type, quantization factors, motion vectors, etc.

The VQM agent then communicates the quality measurements to an aggregator, which sits somewhere in the operator's network. In our demonstration system, the quality is then plotted on a remote device to illustrate the monitoring capability. In a real setting, the values from many devices would be automatically monitored and the aggregate displayed, as is done for more traditional network QoS monitoring.

The VQM agent that is running on the handset is also running on the video server. Thus, the aggregator is receiving initial quality measurements from the server and user quality measurements from the handset. By synchronizing these two streams of data, the aggregator can assess the degradation due to the network. In fact, it is a measure of fidelity that is plotted on the remote device. By reporting fidelity rather than absolute quality, the system can distinguish network induced errors from cases in which the source content is deemed to have low quality to begin with.

For those in the field of objective quality assessment, this sounds like a reduced-reference model rather than a no-reference model. In fact, it may be interpreted as a reduced-reference model using no-reference models as features.

The purpose of the transcoder is as a source of quality degradation. In a demonstration, the bitrate of the transcoder output can be periodically reduced. As this reduction continues, visible compression artifacts will become more and more noticeable. In the demonstration, this degradation in quality is measured by the VQM agent, reported to the aggregator, and reflected in the data plot.

The server and transcoder have also been modified to allow the introduction of packet loss. While the decoder in

the handset will do its best to conceal these errors, visible artifacts can be introduced. Again, the VQM agent estimates the perceptual quality and passes this information up to the aggregator for reporting.

## 5. CONCLUSION

As video becomes a more important and more prevalent type of data on communications networks, new video applications and services are emerging and the need for practical video quality measures is growing. In this paper we identified a number of these new video applications and services. After providing a short survey of the state-of-the-art in no-reference video quality assessment, we presented a sketch and description of a system for demonstrating real-time video quality monitoring.

The system consists of NR video quality measurement agents integrated into video servers and mobile handsets. Both the server and the handset measure the video quality and send reports to an aggregator that calculates a change in quality between the two nodes. This concept is easily extended to multiple nodes as VQM agents are added to the transcoders and other network elements that handle the video as it is transmitted from the source to the handset.

Note also that this system is not limited to mobile handsets. The endpoint could also be a computer receiving streamed video, a set-top box displaying video on a television, or a telepresence or video conferencing device. In all cases, the source, the endpoint, and network elements in between can all be measuring the video quality and sending this data up to the aggregator to allow end-to-end monitoring of video quality. The availability of these measurements will be a key component in enabling the new applications and services discussed in Section 2.

## 6. REFERENCES

- [1] VQEG, "Final report on the validation of reduced-reference and no-reference objective models for standard definition television," 2009.
- [2] A. Bovik, "Handbook of image and video processing", Academic Press, 2000.
- [3] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Applications to JPEG2000", *Signal Proc.: Image Comm.*, vol. 19, pp. 163-172, 2004.
- [4] R. Ferzli, and Lina Karam, "A human visual system-based model for blur/sharpness perception", *VPQM*, 2006
- [5] D. Liu, Z. Chen, F. Xu, and X. Gu "No reference block based blur detection", *QOMEX*, 2009
- [6] Z. Wang, H. Sheikh, and A. Bovik, "No Reference Perceptual Quality Assessment of JPEG Compressed Images", in *IEEE ICIP*, 2002.
- [7] R. Babu, and A. Perkis, "An HVS-Based No-Reference Perceptual Quality Assessment of JPEG Coded Images Using Neural Networks", in *ICIP*, 2005
- [8] Z. Wang, A. Bovik, and B. Evans, "Blind Measurement of Blocking Artifacts in Images", in *IEEE ICIP* 2000.

---

<sup>1</sup> The content was downsized from HD with the goal of maintaining the aspect ratio while downsizing to approximately the same number of macroblocks as CIF.

- [9] R. Muijs, and I. Kirenko, "A No-Reference Blocking Artifact Measure for Adaptive Video Processing", in European Signal Processing Conference, 2005
- [10] J. Y. C. Chen and J. E. Thropp, "Review of Low Frame Rate Effects on Human Performance," IEEE Trans. on Systems, Man and Cybernetics, vol. 37, pp. 1063–1076, Nov. 2007.
- [11] G. Yadavalli, M. Masry, and S. S. Hemami, "Frame Rate Preference in Low Bit Rate Video," in Proc. of ICIP, vol. 1, Nov. 2003, pp. 1–441–4.
- [12] Z. Lu, W. Lin, B. C. Seng, S. Kato, S. Yao, E. Ong, and X. K. Yang, "Measuring the Negative Impact of Frame Dropping on Perceptual Visual Quality," in Proc. SPIE Human Vision and Electronic Imaging, vol. 5666, Jan. 2005, pp. 554–562.
- [13] K.-C. Yang, C. C. Guest, K. El-Maleh, and P. K. Das, "Perceptual Temporal Quality Metric for Compressed Video," IEEE Trans. on Multimedia, vol. 9, pp. 1528–1535, Nov. 2007.
- [14] Y.-F. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling The Impact of Frame Rate on Perceptual Quality of Video," in Proc. of ICIP, San Diego, Oct. 2008, pp. 689 – 692.
- [15] R. R. Pastrana-Vidal and J.-C. Gicquel, "Automatic quality assessment of video fluidity impairments using a no-reference metric", VPQM, 2006.
- [16] S. Wolf, "A no-reference and reduced-reference metric for detecting dropped video frames", VPQM, 2009
- [17] R. Babu, A. Bopardikar, A. Perkis, O. I. Hillestad, "No-Reference metrics for video streaming applications", International Workshop on Packet Video, Dec 2004
- [18] H. Rui, C. Li, S. Qiu, "Evaluation of packet loss impairment on streaming video", J. of Zhejiang University SCIENCE, vol. 7, April 2006.
- [19] A. Reibman, and D. Poole, "Predicting packet-loss visibility using scene characteristics," International Workshop on Packet video, 2007.
- [20] T. Liu, "Perceptual Quality Assessment of Videos Affected by Packet-losses," Ph.D thesis, 2010. Available at: [http://vision.poly.edu/~tliu/Phd\\_Thesis\\_LIU\\_2010.pdf](http://vision.poly.edu/~tliu/Phd_Thesis_LIU_2010.pdf)
- [21] ITU-T Recommendation G.1070, "Opinion model for video-telephony applications", 2007
- [22] VQEG. "Hybrid Perceptual/Bitstream Testplan," 2009.