

This paper appears was presented at the 18th International Symposium on Intelligent Signal Processing and Communications Systems, held December 6-8, 2010 at the University of Electronic Science and Technology of China in Cheng Du, China. The paper appears in the proceedings of that symposium and is available through the IEEE.

© 2010 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

ON QUALITY ASSESSMENT IN THE DELIVERY OF MOBILE VIDEO

Tao Liu, Jeffrey Bloom

Dialogic Research, Inc., Eatontown, NJ
{tao.liu, jeffrey.bloom@dialogic.com}

ABSTRACT

Along with a rapid growing volume of video on communication systems, we see the emergence of many new video applications and services. As video plays a larger role in our communications, the need for accurate, real-time video quality measurement becomes more important for these applications and services.

In this paper, we present an emerging application for quality measurement of transmitted videos, especially delivered for mobile devices, and, in that context, provide a brief survey of the state-of-the-art in no-reference quality models for various kinds of video distortion. We then discuss some challenges that arise when incorporating quality measurement into practical systems.

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

1. INTRODUCTION

The rapid advance of video applications and services is accompanied by an increasing need for accurate, real-time, no-reference video quality assessment. Such tools are needed for many applications including in-network quality monitoring for service level agreement compliance, real-time bandwidth management, control of video quality for tiered services, and quality assurance of end user video.

Objectively measuring video quality will ultimately require models of human vision and models of human visual cognition. An easier task, and more typical approach, is to identify artifacts that are commonly introduced by well defined video processes such as compression, transmission, and resampling in time and space. Even this can be difficult as the artifacts introduced by different transmission components can have dramatically different visual impacts and perceived quality can largely depend on the video content.

In this paper, we present an application of quality assessment in the context of mobile video delivery. We argue that, although any individual measurement of impairment may be insufficient to assess the video quality, the change in impairment measurement as video traverses the network can be used to approximate the degradation in video quality. This is followed by a brief survey of some of the relevant impairment measurement models. Finally, we discuss some of the practical challenges faced in applying existing impairment measures.

2. QUALITY ASSESSMENT IN THE DELIVERY OF MOBILE VIDEO

To illustrate our thoughts about the use of quality measures, we present an emerging application: the use of quality assessment in the delivery of video to mobile devices.

As mobile phones become more advanced and with mobile network capacity increasing, the streaming of video to mobile devices is becoming common. However, the video does not pass directly from the content owner to the mobile device. The video *stream* is handled by a number of different parties before it reaches the viewer. These can include 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) for efficient transport through their specific network. Each such modification has the potential to degrade the video quality. Another major degradation in quality is due to the errors introduced during compression and transmission.

Unfortunately, currently deployed video distribution systems lack good measurement of the video quality throughout the network, particularly where it is most important, at the end-point. There are standard measures of network quality-of-service (QoS): packet loss rate, latency, jitter, etc. But they don't correlate well to visual quality and there are no network QoS tools that can assess the impact of transcoding or transrating on video quality. Instead, video quality measures (VQM) are needed at multiple points in the network as shown in Figure 1. These measures are aggregated and analyzed to provide useful information about the performance of the network from the perspective of video quality.

A number of video service applications can be modeled with a generalized version of Figure 1. Consider the case in which the devices are operated by different companies. At each hand-off point, there are *service level agreements* (SLA) specifying a minimum quality of service. But these SLAs could also specify a maximum amount of degradation to the video quality. With the ability to measure quality, systems could manage their bandwidth usage, insuring that the amount of bandwidth used is just enough necessary to meet the quality targets. Similarly, network operators can establish tiered services in which the video quality delivered to the viewer depends on the price paid. More expensive plans deliver higher quality video. To do this, the quality of the video must be measured and controlled. A final example is quality assurance of end user video. Most video network

operators today are not aware of any video quality problems in their network until they receive a complaint from a customer. A network instrumented to measure video quality will give operators the ability to identify and troubleshoot problems more quickly.

In many cases, the quality measurements shown in Figure 1 can be made with a reference. If the video gateway is modifying the stream, it can measure the quality of the output relative to the input and thus report the level of degradation for which it is responsible. It is not clear, however, how a number of these relative quality measurements can be aggregated to provide insight into the overall impact on quality (it is likely that a simple linear summation would be insufficient). Further, in many applications, the various components in the network are controlled by different parties who each have an incentive to report very slight, if any, degradation in quality; true or not. In addition, there are cases, for example at the mobile device, where there is no reference available.

For these reasons, we propose the use of no-reference models to measure relevant aspects of the video. The quality of the video at the various points in the network is estimated by calculating the change in those measurements as the video passes through the network. This can be considered as a reduced-reference model in which a set of features from the reference is used to assess the quality at the target. The interesting thing here is that we use no-reference measurements as the reduced reference features.

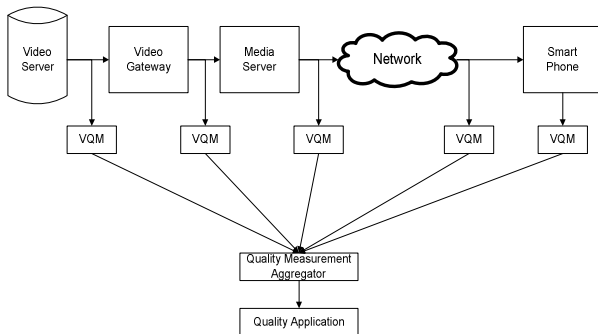


Figure 1 A straw-man network configuration showing how video quality measurement tools might be used to monitor end-to-end changes in video quality.

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

Since the videos to be delivered to mobile devices are generally compressed at relatively low bitrate, low resolution, and low frame rate, and then transmitted via error prone channels, the video quality is subject to some coding-related artifacts and transmission errors. In the remainder of this section, we will discuss no-reference measures for several of these artifacts.

3.1. Compression Artifacts: 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. [1] 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 [2] extended this work by incorporating contrast masking into the edge width calculation. One drawback of the Marziliano et al. algorithm, as recognized in [1] is the reliability of edge detection in the presence of blurriness. For this reason, Liu et al. [3] applied a similar idea, but transition sharpness is measured only at macro-block boundaries, where deblocking filters are applied in H.264 video.

Ferzli and Karam also measure blurriness by characterizing the variance of the statistics impacted by blurriness [2]. Some statistics considered include image variance, gradient, Laplacian, autocorrelation, kurtosis, frequency, and entropy histograms. This approach is simple and does not require edge detection, but the resulting measures have limited accuracy over different video contents.

3.2. Compression Artifacts : Blockiness

Blockiness is introduced during video communications during 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. In the spatial domain, Wang et al. [4] quantified discontinuities by averaging the intensity differences across block vertical and horizontal boundaries. Babu and Perkiş [5] 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. [6] 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 origin 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 [7] where the differences across neighboring lines were accumulated and periodic peaks of these differences were detected as locations of a blockiness grid.

3.3. Packet Loss

Because of network congestion or bit errors due to noise and signal attenuation along the transmission channel, some video data packets may be dropped or discarded. Depending on the compression standard and the resolution of the imagery, a lost packet usually represents the loss of a slice of data or even an entire video 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 and implementation used (error propagation due to packet loss), 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 frame by looking at the pixel data while other work concentrates on assessing the impairment introduced by a known loss.

Pastrana-Vidal and Gicquel [8] detect 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. Wolf [9] detects the loss of frame data which has been concealed by repeating the last frame. This is done by examining the motion energy time history with an adaptive threshold. Spatially, Babu et al. [10] count the number of pixels along the edges where gradient differences are high. Slice Boundary Mismatch (SBM), proposed by Rui et al. [11], considers the strength of continuities as an indication of perceptual distortion. Finally, Reibman and Poole [12] estimate the quality of packet loss impaired video from the perspective of the visibility of packet loss. A generalized linear model is used to leverage a bundle of quality metrics that can be easily extracted or estimated directly from the bitstream.

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 [13], video quality is predicted from a few parameters in the bitstream.

4. PRACTICAL CHALLENGES

As we seek to use some of the existing no-reference models in the context of a system like that, a number of practical challenges arise. In this section we highlight some of those challenges.

4.1. Interpretation of quality scores

As mentioned earlier, many published “quality measures” are actually measures of specific features that are assumed to be introduced by processes that degrade the visual quality. Examples given above included measures of blockiness and blurriness. However, associating these features with quality can be misleading because there are examples of high quality, non-impaired imagery that would be deemed blocky or blurry. Consider a close-up photograph of a bug on a leaf in which the majority of the image may be out of focus.

In many practical situations, the interesting measure is not really the absolute blurriness or blockiness, but rather the change in that measure across the network. A user-generated video, captured by a cell phone and uploaded to a video sharing site may have low quality. But the network operator is only interested in the impact of the network on quality, not the original quality.

Both of these issues can be addressed by using no-reference measures at the video source and at the video endpoint, but instead of interpreting these measurements as an indication of quality, the difference in these measurements is used as an indication of the change in quality. Thus, the blurry scene of a bug on a leaf is still deemed blurry, but a small change in blurriness from the server to the phone indicates that the network delivered a high fidelity copy.

The system shown in Figure 1 takes this one step further. By accessing measurements at multiple points in the network, the system can identify the point in the network that is introducing degradations.

4.2. Fully decoding vs. partial decoding

The ITU-T recommendation G.1070 is an example of video quality model based on information obtained from a compressed bitstream. At the other extreme, the blurriness measures mentioned in Sec. 3 rely entirely on an analysis of the decoded pixel data. It is generally believed that better predictions about quality can be obtained by incorporating as much information as possible and new methods are emerging that make use of both bitstream and pixel domain analysis.

However, the amount of information available to a no-reference quality prediction tool varies depending on its location within the network. Measuring quality at the end point can analyze the bitstream prior (or during) decoding and can theoretically access the pixel data after decoding just prior to display. Similarly, since conceptually a transcoder decodes and then re-encodes the video, a video

quality prediction tool at a transcoder could access the data in both domains. In contrast, there are other network nodes that do not decode the video. Consider measurement of the video quality at a video server, gateway or router which may process hundreds or thousands of video streams at the same time. Depending on the computational resources available (see Sec. 4.4) a full decoding for the purpose of assessing the quality of the pixels may not be possible. Instead, those measurement tools may need to rely on partial decoding, to access motion vectors, DCT coefficients, etc.

4.3. Consistency of quality metrics at multi-points

The two problems discussed above combine to create an additional challenge. If no-reference quality measurements are collected at multiple points in the network and then compared to assess the change in quality, then the measurement tools at all points in the network should be the same. Then, since different points in the network have access to different representations of the data, the common measurement tool must be the most restrictive. In other words, the set of data used as input to the measurement tool should be the common set available at all points in the network. This typically restricts the measurement tool to strictly bitstream models.

Unfortunately, as implied above, these models are not as accurate at predicting perceptual quality as those that have access deeper into the partially decoded or fully decoded data. Additionally, there is the fact that compression standards typically do not specify how a decoder should respond to errors in the bitstream. Different decoders will produce different decoded pixels in the presence of bitstream errors. Thus, the visual impact of those bitstream errors cannot be accurately predicted by analyzing the bitstream alone.

4.4. Computational constraints

The consistency challenge discussed above is primarily due to tight computational constraints at some measurement points. An obvious example is measurement in a mobile device. These devices have limited available resources including battery power, memory, and compute cycles. However, computational challenges exist in less likely spots as well. A video server may have very powerful processors, large memory footprints, and plenty of electrical power, but these devices are also tasked with serving 1000's of streams simultaneously. Adding a full decoder and quality measurement tool to each stream becomes unreasonable.

4.5. Synchronization issues

There are two synchronization issues that arise in the implementation of a system similar to that shown in . First, consider multiple network devices (many versions of server, network, end-point all running in parallel), all reporting quality measurements to a single aggregator. The system

must be able to establish which measurements can serve as references to which other target measurements.

Once that first synchronization issue has been addressed, the two streams of measurement data, target and reference, must be temporally aligned.

5. CONCLUSION

In this paper, we presented a straw-man network configuration showing how no-reference video quality measures can be used effectively. After providing a short survey of the state-of-the-art in no-reference video quality assessment, we identified a number of engineering challenges that arise in the application of those tools to a practical system.

One main point of this paper is that while current no-reference models may be insufficient in their ability of directly predicting subjective quality, they may still be useful for measuring the change in quality across a network. A second conclusion is that there is a gap in the research of quality assessment tools to measure for impairments introduced during lossy compression as well as for impairments that result from transmission errors (modeled as packet loss).

6. REFERENCES

- [1] 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.
- [2] R. Ferzli, and Lina Karam, "A human visual system-based model for blur/sharpness perception", *VPQM*, 2006
- [3] D. Liu, Z. Chen, F. Xu, and X. Gu "No reference block based blur detection", *QOMEX*, 2009
- [4] Z. Wang, H. Sheikh, and A. Bovik, "No Reference Perceptual Quality Assessment of JPEG Compressed Images", in *IEEE ICIP*, 2002.
- [5] R. Babu, and A. Perkis, "An HVS-Based No-Reference Perceptual Quality Assessment of JPEG Coded Images Using Neural Networks", in *ICIP*, 2005
- [6] Z. Wang, A. Bovik, and B. Evans, "Blind Measurement of Blocking Artifacts in Images", in *IEEE ICIP* 2000.
- [7] R. Muijs, and I. Kirenko, "A No-Reference Blocking Artifact Measure for Adaptive Video Processing", in 'European Signal Processing Conference' 2005
- [8] R. R. Pastrana-Vidal and J.-C. Gicquel, "Automatic quality assessment of video fluidity impairments using a no-reference metric", *VPQM*, 2006.
- [9] S. Wolf, "A no-reference and reduced-reference metric for detecting dropped video frames", *VPQM*, 2009
- [10] R. Babu, A. Bopardikar, A. Perkis, O. I. Hillestad, "No-Reference metrics for video streaming applications", *International Workshop on Packet Video*, Dec 2004
- [11] H. Rui, C. Li, and S. Qiu, "Evaluation of packet loss impairment on streaming video", *J. of Zhejiang University SCIENCE*, vol. 7, April 2006.
- [12] A. Reibman, and D. Poole, "Predicting packet-loss visibility using scene characteristics," *International Workshop on Packet video*, 2007.
- [13] ITU-T Recommendation G.1070, "Opinion model for video-telephony applications", 2007