Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications

Stefan Winkler, Animesh Sharma, David McNally

Genimedia, EPFL – PSE
1015 Lausanne, Switzerland
[swinkler@genimedia.com]

Abstract

Guaranteeing a certain level of quality for multimedia streaming applications is quite well understood in terms of network QoS, but it is much more tenuous in terms of perceptual quality as perceived by the user. In this paper, we classify video quality measurement schemes and review existing approaches with a focus on non-intrusive quality metrics, which do not require access to the reference video. In particular, we evaluate three different no-reference blockiness metrics and compare their performance.

Keywords

Perceptual QoS, visual quality assessment, blocking artifact, no-reference blockiness metrics

1. Introduction

The development of powerful compression algorithms and the rapid growth of network resources have facilitated the widespread distribution of video and multimedia in digital form. Keeping the bandwidth and storage requirements to a minimum while maintaining good visual quality has been the priority in the design of new digital video systems, and guaranteeing a certain level of quality is an important concern for content providers.

Variations in quality are due to lossy compression as well as transmission errors, which both lead to artifacts in the received material. The amount and visibility of these distortions strongly depend on the actual video content. The accurate measurement of quality as perceived by the user has become one of the great challenges.

Naturally, the benchmark for any kind of video quality assessment are subjective experiments, where a number of people are asked to watch test clips and to rate their quality. Several procedures for such experiments have been formalized in ITU-R Recommendation BT.500 [2], which suggests standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods. The problem with subjective experiments is that they they are time-consuming, hence expensive and often impractical. Furthermore, for many applications (e.g. online quality monitoring and control) subjective experiments cannot be used at all.

Given these limitations, engineers have turned to simple error measures such as mean squared error (MSE) or peak signal-to-noise ratio (PSNR), suggesting that they would be equally valid. However, these simple measures operate solely on the basis of pixel-wise differences and neglect the imact of video content and viewing conditions on the actual visibility of artifacts. Therefore, their predictions often do not agree well with perceived quality.

Another way to perform objective measurements of data transmission is looking at bit error rate (BER), packet loss ratio (PLR) and other network-related parameters. Establishing and maintaining a certain level of network quality of service (QoS) for different applications is a very active research area at the moment [5, 12], but again the measurements and protocols used there are oblivious to the actual content being transmitted over the network and have no direct relation to the video quality as perceived by the user.

2. Video Quality Metrics

The shortcomings of these methods led to the study of more advanced perceptual quality metrics in recent years. An up-to-date review of such metrics can be found in [17]. In principle, two different approaches can be distinguished:

Approaches based on models of the human visual system are the most general and potentially most accurate ones [15]. Examples of such metrics are described in [4, 9, 16] among others. However, the human visual system is extremely complex, and many of its properties are not well understood even today. Besides, implementing these models is computationally expensive due to their complexity.

On the other hand, metrics need not necessarily rely on general models of the human visual system; they can exploit a priori knowledge about the compression and transmission methods as well as the pertinent types of artifacts using ad-hoc techniques or simple specialized vision models. Examples of such specialized metrics include [8, 14]. While such metrics are not as versatile, they normally perform well in a given application area. Their main advantage lies in the fact that they often permit a computationally more efficient implementation.

Several of these video quality metrics were compared against subjective ratings for a well-defined set of test sequences in an ambitious performance evaluation undertaken by the Video Quality Experts Group (VQEG). The work and findings of VQEG are described in [7] and in the group's final report [11] in more detail. Consult VQEG's web site `http://www.crc.ca/vqeg/` for an overview of its current activities.

2.1. Out-of-Service vs. In-Service Metrics

The emphasis of most metrics today is *out-of-service* testing (see Figure 2), where the full reference video is available to the metrics. This is quite a severe restriction on the kind of applications such a metric can be used for, however.

Figure 1: Non-intrusive/no-reference (left) and intrusive/reduced-reference (right) in-service testing setup.

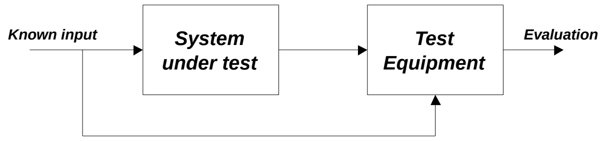


Figure 2: Out-of-service testing of a system.

In-service metrics, which are designed to monitor and control systems while they are in operation, are much more powerful. They can be used to carry out measurements at practically any point of the transmission chain. This is a particularly important issue in multimedia streaming applications. The setup can be intrusive or not, depending on the objective of the test and the nature of the testing methodology; Figure 1 illustrates both cases. The corresponding metrics are often referred to as reduced-reference and no-reference metrics, respectively.

The fact that the full reference video is usually not available for comparison makes an accurate assessment much more difficult for in-service metrics. Therefore, the algorithms are generally based on some a priori knowledge about the scene content, the compression method and/or the expected artifacts. Several methods have been proposed recently [1, 19]. Most of them aim at identifying certain features in a scene and assessing their distortion. They also take into account knowledge of the compression method and the corresponding artifacts (a detailed overview of typical video compression artifacts and their causes can be found in [22]).

3. Blockiness Metrics

One of the most common artifacts in compressed video is blockiness. Blockiness manifests itself as the appearance of a block structure in the video. It is caused by block-based coding schemes such as H.261, H.263, MPEG-1, MPEG-2 and MPEG-4, which compute the DCT on every 8×8 block in the image and quantize each block's coefficients separately.

A number of blockiness metrics have been proposed [3, 6, 21], but these techniques require access to the reference image or video. We have implemented three no-reference blockiness metrics recently proposed in the literature:

3.1. Vlachos Metric

Vlachos [10] uses an algorithm based on the cross-correlation of subsampled images. The sampling structure is chosen such that every sub-image contains one specific pixel from each 8×8 block. Four sub-images are constructed from the four corner pixels of each block. Four more sub-images are constructed from four neighboring pixels in the top left corner of each block. Finally, the cross-correlations among the former four sub-images are normalized by the cross-correlations among the latter four sub-images to yield a measure of blockiness.

3.2. Wang-Bovik-Evans Metric

Wang, Bovik and Evans [13] model the blocky image as a non-blocky image interfered with a pure blocky signal. They apply 1-D FFTs to horizontal and vertical difference signals or rows and columns in the image to estimate the average horizontal and vertical power spectra. Peaks in these spectra due to 8×8 block structures are identified by their locations in the spectra. The power spectra of the underlying non-blocky images are approximated by median-filtering these curves. The overall blockiness measure is then computed as the difference between these power spectra at the locations of the peaks. The integration of visual masking effects is briefly described as well.

3.3. Wu-Yuen Metric

Wu and Yuen [20] measure the horizontal and vertical differences between the columns and rows at all 8×8 block boundaries. Weights for taking into account perceptual luminance- and texture-masking effects are derived from the means and standard deviations of the blocks adjacent to each boundary. The resulting measure is normalized by an average of the same measures computed at non-boundary columns and rows.

3.4. Experiments and Discussion

To verify and compare these blockiness metrics, test clips described in [18] were used. The test scenes are taken from the VQEG set [11] and are 8 seconds long with a frame rate of 25 Hz. They were de-interlaced and subsampled to a resolution of 360×288 pixels per frame for progressive display on a computer screen. The Microsoft MPEG-4 codec (version 2) at 1 Mb/s and the Sorenson Video codec (version 2.11) at 2 Mb/s were used to create the test sequences. Sample frames from the ”racecar” clip are shown in Figure 3.

The results produced by our implementation of the three above-mentioned blockiness metrics are shown in Figure 4. We have run the algorithms also on the uncompressed clip for comparison purposes. Perceptually, the test sequence starts out practically devoid of blocking artifacts; blocks start appearing about halfway through, peaking soon after and decreasing slightly towards the end. The uncompressed clip exhibits no blockiness at all. The evolution over time predicted by the blockiness metrics thus corresponds quite well to perceived blockiness.

On the other hand, two of the three algorithms rate the blockiness produced by the Sorenson codec significantly lower than the MPEG-4 codec, while visually the opposite is true. An explanation for this is that the extreme blocking artifacts produced by the Sorenson codec remove many of the intensity differences at block boundaries which the algorithms rely on. Furthermore, it adds many blocks of size 16×16, while the algorithms are tuned for 8×8 blocks. The sharp drops at frame numbers 125, 150 and 175 for this sequence are due to the keyframe interval of 1 second, and an inspection of these frames reveals that they really exhibit no blocking artifacts.

Vlachos' metric shows the least distinction between the uncompressed and the compressed sequences. However, it must be noted that we chose to implement the cross-correlations in a different manner, which may be responsible for this below-par prediction performance.

One common drawback of all three blockiness metrics is that the exact location of block boundaries must be known. Spatial shifts of the block structure with respect to the origin are assumed to be zero, which may not always be the case. Likewise, blocks not aligned with the 8×8 grid will not be recognized. Furthermore, none of the metrics are explicitly targeted at video, as they process each frame separately. Thus, the effect of motion on block visibility in video is not taken into account.

4. Conclusions

Perceptual quality measurement has become a very active area of research. Full-reference out-of-service metrics are rather well established; current efforts focus on reduced- and no-reference metrics, which are required for in-service measurement, monitoring and control. While most existing video quality metrics still focus on television production and broadcast, some are beginning to target multimedia applications, which are much less constrained. The algorithms and metrics reviewed in this paper represent important steps towards comprehensive no-reference video quality metrics.

5. References

[1] P. Bretillon, J. Baina: "Method for image quality monitoring on digital television networks." in *Proc. SPIE*, vol. 3845, Boston, MA, 1999.

[2] ITU-R Recommendation BT.500-10: "Methodology for the subjective assessment of the quality of television pictures." ITU, Geneva, Switzerland, 2000.

[3] S. A. Karunasekera, N. G. Kingsbury: "A distortion measure for blocking artifacts in images based on human visual sensitivity." *IEEE Trans. Image Processing* **4**(6):713–724, 1995.

[4] J. Lubin, D. Fibush: "Sarnoff JND vision model." T1A1.5 Working Group Document #97-612, ANSI T1 Standards Committee, 1997.

[5] A. Mishra: *Quality of Service in Communications Networks*. John Wiley & Sons, 2001.

[6] M. Miyahara, K. Kotani: "Block distortion in orthogonal transform coding – analysis, minimization and distortion measure." *IEEE Trans. Comm.* **33**(1):90–96, 1985.

[7] A. M. Rohaly et al.: "Video Quality Experts Group: Current results and future directions." in *Proc. SPIE*, vol. 4067, pp. 742–753, Perth, Australia, 2000.

[8] K. T. Tan, M. Ghanbari, D. E. Pearson: "An objective measurement tool for MPEG video quality." *Signal Processing* **70**(3):279–294, 1998.

[9] C. J. van den Branden Lambrecht: *Perceptual Models and Architectures for Video Coding Applications*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1996.

[10] T. Vlachos: "Detection of blocking artifacts in compressed video." *Electronics Letters* **36**(13):1106–1108, 2000.

[11] VQEG: "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment." 2000, available at `http://www.crc.ca/vqeg/`.

[12] Z. Wang: *Internet QoS: Architectures and Mechanisms for Quality of Service*. Morgan Kaufmann, 2001.

[13] Z. Wang, A. C. Bovik, B. L. Evans: "Blind measurement of blocking artifacts in images." in *Proc. ICIP*, vol. 3, pp. 981–984, Vancouver, Canada, 2000.

[14] A. B. Watson et al.: "Design and performance of a digital video quality metric." in *Proc. SPIE*, vol. 3644, pp. 168–174, San Jose, CA, 1999.

[15] S. Winkler: "Issues in vision modeling for perceptual video quality assessment." *Signal Processing* **78**(2):231–252, 1999.

[16] S. Winkler: "A perceptual distortion metric for digital color video." in *Proc. SPIE*, vol. 3644, pp. 175–184, San Jose, CA, 1999.

[17] S. Winkler: *Vision Models and Quality Metrics for Image Processing Applications*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2000.

[18] S. Winkler: "Visual fidelity and perceived quality: Towards comprehensive metrics." in *Proc. SPIE*, vol. 4299, San Jose, CA, 2001.

[19] S. Wolf, M. H. Pinson: "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system." in *Proc. SPIE*, vol. 3845, pp. 266–277, Boston, MA, 1999.

[20] H. R. Wu, M. Yuen: "A generalized block-edge impairment metric for video coding." *IEEE Signal Processing Letters* **4**(11):317–320, 1997.

[21] Z. Yu et al.: "Vision model based impairment metric to evaluate blocking artifacts in digital video." *Proc. IEEE* 2001, to appear.

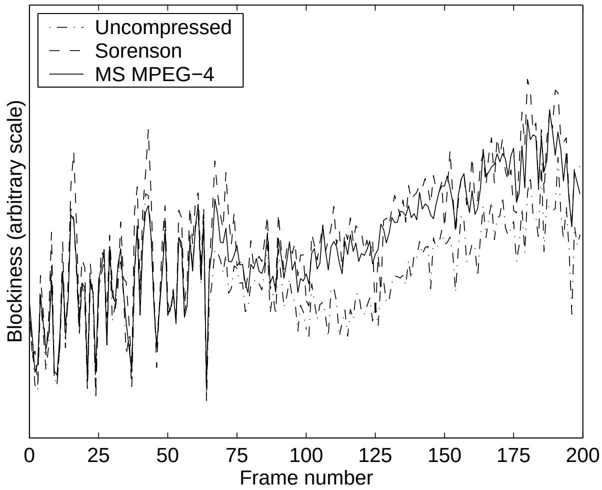
[22] M. Yuen, H. R. Wu: "A survey of hybrid MC/DPCM/DCT video coding distortions." *Signal Processing* **70**(3):247–278, 1998.

(a) Uncompressed

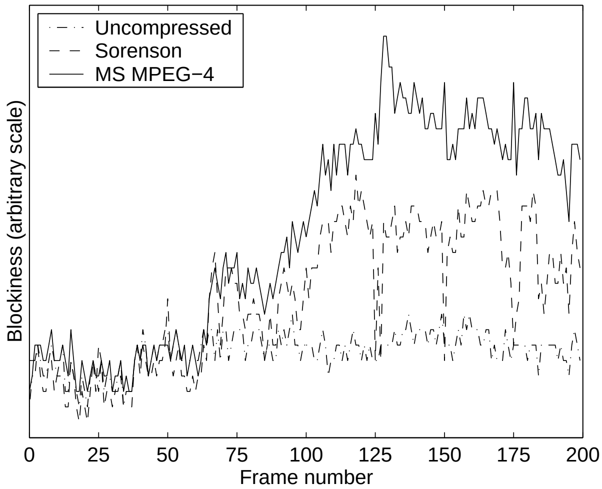
(b) Microsoft MPEG-4 codec

(c) Sorenson Video codec

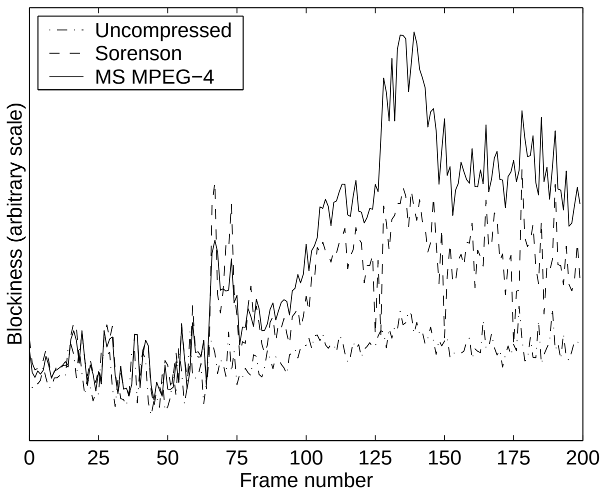
Figure 3: Frame from ”racecar” test clip.



(a) Vlachos metric [10]



(b) Wang-Bovik-Evans metric [13]



(c) Wu-Yuen metric [20]

Figure 4: Comparison of blockiness metrics.