VIDEO QUALITY EVALUATION FOR UMA

Odd Inge Hillestad, R. Venkatesh Babu, Ajit S. Bopardikar, Andrew Perkis

Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology
Trondheim, Norway

ABSTRACT

This paper[1] deals with monitoring user perception of multimedia presentations in a Universal Multimedia Access (UMA) enabled system using objective metrics. To this end, we propose a no-reference (NR) video quality metric that measures the block-edge impairment or blockiness in a received video sequence. This NR metric is designed to be used as an integral part of the UMA viewer at the user end. It is based on counting the number of blocks in a given frame that would contribute to the overall perception of blockiness for that frame. It is based on the idea that block-edge impairment is observed in regions with low spatial activity. Our metric increases as compression increases while at the same time remaining close to zero when monitoring an uncompressed original sequence with no blockiness. The metric has low computational complexity and can be used for real-time monitoring of streaming video in a multimedia transmission scenario.

1. INTRODUCTION

An increasing demand for ubiquitous access to multimedia content, and a corresponding increase in the variety and amount of content being produced, end-user terminals and networking facilities, calls for a solution which can facilitate a good user experience of media consumption. Some essential aspects of this problem are being addressed through the concept of Universal Multimedia Access (UMA) [1] which deals with the delivery of images, video, audio and multimedia content in general under various network access and resource conditions, communication device capabilities and end user preferences. The objective of UMA enabled systems is to provide the user with the best possible subset of a multimedia resource that the user is capable of receiving. In this sense, the concept of UMA deals with quality with respect to the delivery of content. The quality is treated as an end-to-end Quality of Service aggregate which we choose to view as *Quality of Experience* (QoE). Increasingly, this idea is evolving to include the User and the User's perception of the media being delivered. In this premise, known as the Universal Multimedia Experience (UME)[2], the network and the terminal are considered purely as means to deliver the content. The aim of this paradigm shift is to enable adaptation of the media content presented to the end User based on that User's perception of that content in a specific environment and context. In other words, UME emphasis the end user, and the ultimate goal is to provide the end user with meaningful content that maximizes the user's (QoE).

[1]This work was supported by the Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence" appointed by The Research Council of Norway. http://www.ntnu.no/Q2S/

A generic UMA-enabled communication device used to consume a multimedia presentation is called a UMA viewer and is central to obtain the notion of QoE. In addition to being a media player, it is required that the UMA viewer incorporate an awareness of UME, resulting in an intelligent behavior regarding how the content is presented, delivered, and ultimately, how the media is perceived by the end user. The latter is a subjective attribute that depends on several sensory factors that are not completely understood and are difficult to evaluate. Nonetheless, there is a clear need for automated evaluation of perceived quality of the rendered multimedia content. This means, we require a metric that will give us a measure for the quality of the rendered content that is strongly correlated with how the content is perceived by a cross section of the users. In general, such a metric would have to satisfy certain conditions. For one, the quality metric would have to have a low computational complexity. It would also be required to perform consistently over a wide range of content types. In many situations such as streaming of video one would require a metric that could evaluate the perceptual quality of the content with either limited or no access to reference content. Such metrics are called reduced-reference (RR) and no-reference (NR) metrics, respectively [3, 4]. Metrics that estimate the perceived quality using the uncompressed original as reference, are called full-reference (FR) metrics.

Video quality metrics is currently a research area gaining an increasing amount of attention. In this paper, we deal with NR video metrics for a UMA viewing environment. In particular, we present a model for a UMA capable viewer and introduce a novel NR metric that can be used in its framework.

The paper is organized as follows: Section 2 describes the UMA viewer requirements. Section 3 discusses our proposed video metric. The results and discussion are presented in Section 4, and Section 5 concludes the paper.

2. REQUIREMENTS FOR THE UMA VIEWER

The UMA concept places certain requirements on the viewers used for consumption of the presentation. Some of these can be summarized for traditional multimedia consumption as follows:

- Being able to buffer a given amount of data to prevent frame delays during small network traffic variations when the channel characteristics change dynamically. Commercial viewers use this today; the problem is how to take control of the buffer based on dynamic channel and network feedback.
- Being able to use a media description annotation to automatically extract media conversions from an original sequence. This should be done instantaneously and continuously, or alternatively the media descriptor can be as simple

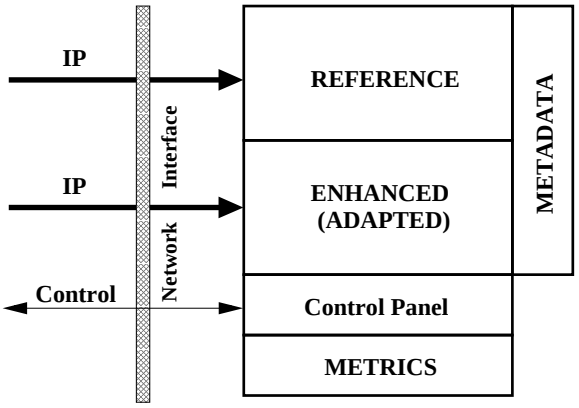

Fig. 1. The UMA Viewer.

as a pointer to the correct conversion of the content on the server.

- Being able to provide adequate support for negotiation procedures during time variations in channel conditions and access schemes (also for initial setup).
- Being able to provide intelligent QoS control of the streaming of the content, including fast response to changes in channel bandwidth and automated presentation of changing frame rates.
- Being able to support mechanisms and metrics for quality evaluation and monitoring in order to provide maximum QoE.

Packet switched communication using the Internet protocol (IP) is becoming the common denominator for rapidly growing areas of multimedia services and wireless access. Multimedia over IP and wireless networks face many challenges due to the intrinsic natures of these networks such as unknown and dynamic bandwidth, delay jitter and packet-loss. This imposes some necessary trade-offs between QoS guarantee and resource utilization efficiency. These problems need to be tackled intelligently for efficient delivery of multimedia content for various Users.

At present there is no viewer which can provide all the above mentioned basic functionalities. It is therefore necessary to design a viewer incorporating all these functionalities. As a first attempt we consider the above requirements and propose an architecture as shown in Fig. 1. This architecture is meant for laboratory evaluation of an UMA viewer. Here the 'reference' is the original streamed media and the 'enhanced' is the stream adapted to suit the present network conditions, channel variations and user capabilities. The 'reference' is used in computing 'Full-reference' (FR) and 'Reduced-reference' (RR) metrics. The enhanced presentation makes use of the metrics measuring quality using NR methods.

The 'control' is the feedback signal to the streaming server, which requests for adaptation of the media. The 'metrics' window provides the information regarding the quality of the adapted content. Finally the 'metadata' window lets the viewer show and use available content and content descriptors.

In the following section we propose a new NR metric for measuring the quality of compressed video.

3. A NO-REFERENCE BLOCKINESS METRIC FOR VIDEO

As mentioned above, NR metrics are useful in scenarios where access to the reference video stream is not available. With no reference to compare with, NR metrics attempt to quantify the effects of various distortion artifacts on the perception of the content. In particular, for block-based video compression schemes such as the MPEG and ITU standards (e.g. MPEG-1/2/4, H.263/4), the main forms of distortions include blocking effect, blurring, ringing and the DCT basis image effect[5, 6]. NR metrics proposed have usually tried to quantify the effects of these distortions [7, 8] but the emphasis of research on NR metrics has been predominantly on quantifying the effects of blocking artifacts[9, 4, 10, 11]. This is because blocking artifacts tend to be perceptually the most significant of all coding artifacts[9]. With the Video Quality Experts Group (VQEG) working towards their standardization, NR metrics remain a topic of great research interest.

Most algorithms that measure block-edge impairment make use of the fact that block-edge gradients can be masked because of spatial activity around them (spatial or texture masking), or may not be discernible in very dark or bright regions [4, 9, 11]. Block-edge gradients are typically computed as a function of the abrupt change in pixel values across a horizontal or vertical block-edge. Spatial activity is the degree of variation in pixel values in an area of the image, for instance the variation inside a block or near a block boundary. The higher the variation, the higher the spatial activity and better is its capacity to mask block-edge impairment. Thus, ideally, an NR blockiness metrics should measure the users perception of blockiness in each video frame and do so with low computational complexity so that it can be used for real-time monitoring. In the next subsection we describe a novel low-complexity blockiness metric based on the ideas mentioned above.

3.1. The proposed blockiness metric

The metric proposed in this work is based on the idea that a block-edge gradient can be masked by a region of high spatial activity around it. It can be observed that blockiness perceived in a frame is usually because of blocks with at least one edge exhibiting low activity. Let B_{ij} represent an 8×8 block of pixels starting at location (i, j) in a given frame. I_k,$k = 1, \ldots, 4$, represents the edges of the block as shown in Figure 2.

To measure the activity along a given edge I_k we first divide it into three segments of length 6, namely, a_{k1}, a_{k2} and a_{k3}.

$$\begin{aligned} a_{k1} &= I_k(n) : n = 0 \ldots, 5 \\ a_{k2} &= I_k(n) : n = 1 \ldots, 6 \\ a_{k3} &= I_k(n) : n = 2 \ldots, 7 \end{aligned} \quad (1)$$

This is shown in Figure 3. We define activity as the standard deviation, σ_{kl} for each a_{kl}, and $l = 1, \ldots, 3$. For a given edge I_k, we define the activity to be low if at least one of σ_{kl}, $l = 1, \ldots, 3$, is below a chosen threshold ε. In other words, if there is at least one segment of the edge which has low activity (standard deviation) then the edge and thus the block it belongs to can contribute to the overall perception of blockiness of the frame.

The metric is then computed as follows. For each frame:

1. initialize the block counter $C_B = 0$.
2. In each block B_{ij} along each edge I_k, for each a_{kl}, $k = 1, \ldots, 4$ and $l = 1, \ldots, 3$ compute the standard deviation,

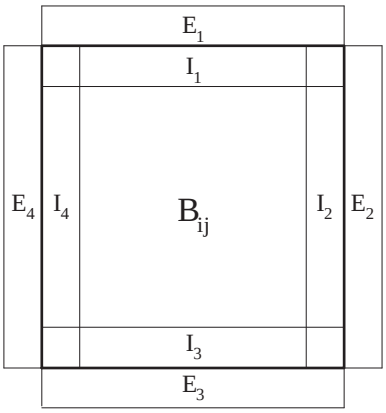

Fig. 2. An 8×8 block and its edges.

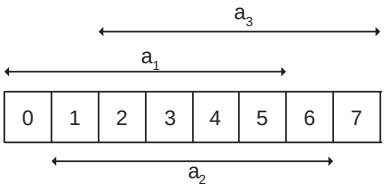

Fig. 3. An 8×8 block and its edges.

σ_{kl}.Thus we obtain three activity measures per edge giving us a total of twelve activity measures.

3. Now compute the gradient corresponding to each a_{kl}

$$\begin{aligned} \Delta_{k1} &= \text{mean}|I_k(n) - E_k(n)| : n = 0\ldots,5 \quad (2) \\ \Delta_{k2} &= \text{mean}|I_k(n) - E_k(n)| : n = 1\ldots,6 \\ \Delta_{k3} &= \text{mean}|I_k(n) - E_k(n)| : n = 2\ldots,7 \end{aligned}$$

where E_k, $k = 1, \ldots, 4$ are the edges adjacent to the corresponding block edges, E_k, $k = 1, \ldots, 4$, as shown in Figure 2.

4. If at least one segment satisfies

$$\begin{aligned} \sigma_{kl} &< \varepsilon \quad (3) \\ \Delta_{kl} &> \tau \end{aligned}$$

$k = 1, \ldots, 4$ and $l = 1, \ldots, 3$, increment C_B by 1. That is, we count B_{ij} as contributing towards the overall perception of blockiness of the frame.

The overall blockiness measure $\mathcal{B}_F$ for the present frame, is then

$$\mathcal{B}_F = \frac{C_B}{\text{Total number of blocks in the frame}}. \quad (4)$$

Clearly, the range of the metric is $[0, 1]$ where 0 corresponds to no blockiness and 1 to the scenario where all the blocks in a frame are visible. The bit depth for the video sequence is assumed to be 8 bits or 255 grayscale levels. The value of ε is chosen as a threshold to isolate edges with low activity. To this end we chose $\varepsilon = 0.1$. This corresponds to the situation when there is a minimal deviation from the mean of the segment. Increasing the value of ε would result in edges with a greater standard deviation being picked. This increases the possibility of counting blocks with segments that might have enough spatial activity to mask the block-edge gradient for that edge.

The value of τ can be chosen so that given low activity, the largest number of perceivable block-impaired edges will be counted in the metric. Increasing the value of τ would mean rejecting segments with low spatial activity which also have a block-edge gradient that can be perceived. On the other hand, choosing a very small value of τ would result in a situation where an imperceptible edge might result in a block being counted, thus giving a false reading. For our simulations, we chose a value of $\tau = 2.0$ because we found that this value of τ gave us the best performance a for wide range of video sequences.

4. RESULTS AND DISCUSSION

For our simulations we considered 10 sec. video sequences in CIF resolution (frame size of 352×288), 30 frames/sec and YUV (4:2:0) format. For results presented here we only consider the Y or the luminance channel. The original video sequence was encoded at various bitrates using the XviD MPEG-4 ASP codec [12] with a GOP size of 30 frames. We compare the performance of the proposed metric with the Wang, Sheik and Bovik (WSB) quality assessment model[10]. MATLAB code for the model was obtained from [13]. Because the WSB metric increases with image quality, and typically has range of 0 to 10, we normalize by 10 and subtract the result from 1. This procedure allows us to compare its performance with that of the proposed metric.

Both metrics were computed for each frame of the original and the encoded sequences. Here we present results obtained for one specific sequence, namely, the "Paris" sequence. Figure 4 shows the result of applying the proposed NR metric to the first two GOPs (frames 1-60) of this sequence and Figure 5 shows the corresponding results for the WSB metric. Note that the proposed metric is nearly zero for the original sequence. In other words, it measures no blockiness in the uncompressed original video as expected. We also see that both metrics increase as the compression increases or equivalently, the bit rate decreases. This is in keeping with the fact that higher compression implies coarser quantization and consequently increased perceived blockiness. The peaks in both figures indicate the *I* (intracoded) frame, and suggest that blockiness perceived in the I-frame is the highest in a GOP at all bit rates. This can also be verified by visual inspection.

Figure 6 shows the change in both metrics for one frame, namely, frame number 31 which is an I (intracoded) frame encoded at different rates, namely, 1234 Mbps, 699 kbps, 489 kbps, 346 kbps, 233 kbps, 186 kbps 147 kbps and 128 kbps. It can be seen that both curves show a graceful degradation. Thus, the proposed metric compares favorably with the WSB metric.

5. CONCLUSION

In this paper, we propose a novel No-Reference metric intended to measure the blocking artifacts in compressed video and present examples of its performance. The metric can be used as an integral part of a complete UMA viewer as a real-time monitoring tool. It could also be used in other applications such as post processing video frames for improved perceptual quality.

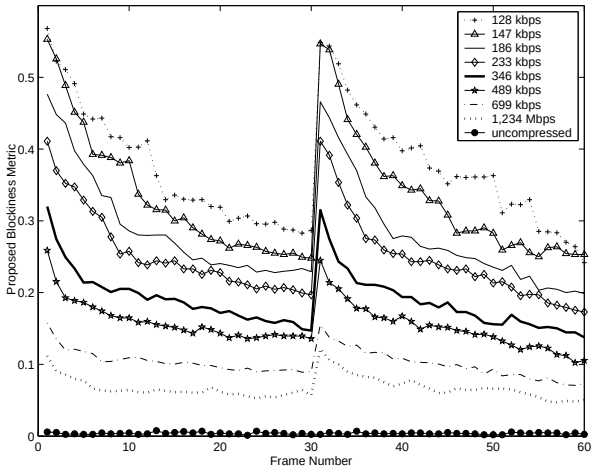

Fig. 4. Proposed blockiness metric for the first 30 frames of the "Paris" sequence coded at different bitrates.

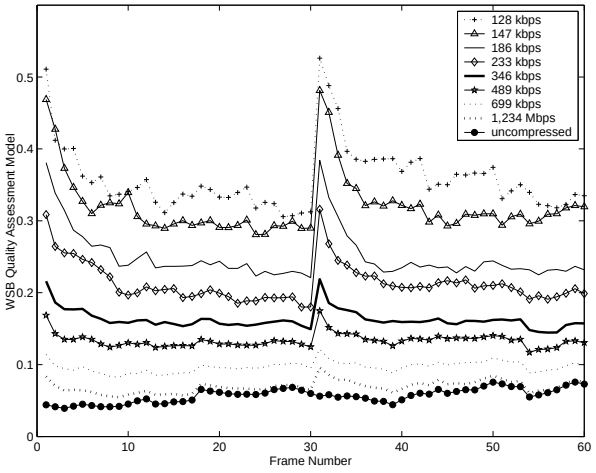

Fig. 5. WSB metric for the first 30 frames of the "Paris" sequence coded at different bitrates.

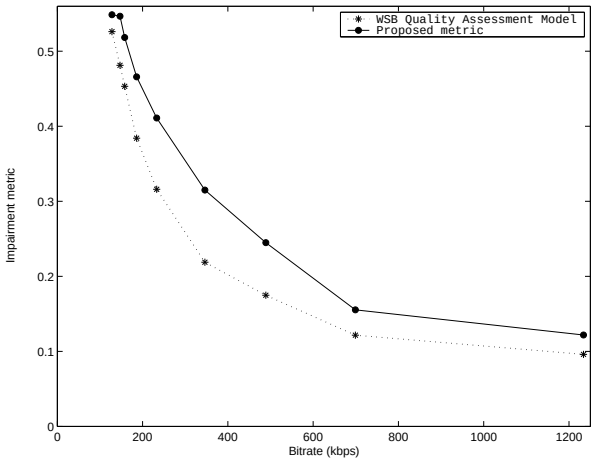

Fig. 6. Comparison of the proposed metric and the WSB metric for frame 31 of the "Paris" sequence at different bit-rates.

6. REFERENCES

[1] A. Perkis, Y. Abdejaoued, C. Christopoulos, T. Ebrahimi, and J. F. Chicharo, "Universal multimedia access from wired and wireless systems," *Circuits, Systems and Signal Processing; Special issue on Multimedia Communications*, vol. 20, no. 3, pp. 387–402, 2001.

[2] F. Pereira and I. Burnett, "Universal multimedia experiences for tomorrow," *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 63–73, March 2003.

[3] L. Lu, Z. Wang, A. C. Bovik, and J. Kouloheris, "Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of mpeg video," in *IEEE International Conference on Multimedia and Expo*, Yorktown Heights, NY, US, 2002, pp. 61– 64.

[4] S. Winkler and A. Sharma and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in *Proc. 4th International Symposium on Wireless Personal Multimedia Communications*, Aalborg, Denmark, September 2001, pp. 553–556.

[5] H. R. Wu, M. Yuen, and B. Qiu, "Video coding distortion classification and quantitative impairment metrics," in *International Conference on Signal Processing*, October 1996, vol. 2, pp. 962–965.

[6] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 4, no. 11, pp. 317–320, November 1997.

[7] J. Caviedes and S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," in *Proceedings of the International Conference on Image Processing*, Rochester, NY, September 22–25, 2002, vol. 3, pp. 53–56.

[8] Pina Marziliano, Frédéric Dufaux, Stefan Winkler, and Touradj Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings of the International Conference on Image Processing*, Rochester, NY, September 22–25, 2002, vol. 3, pp. 57–60.

[9] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 70, no. 3, pp. 247–278, November 1998.

[10] Z. Wang and H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. ICIP'02*, September 2002, vol. 1, pp. 477–480.

[11] W. Gao, C. Mermer, and Y. Kim, "A de-blocking algorithm and a blockiness metric for highly compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1150–1159, December 2002.

[12] "Website: http://www.xvid.org," .

[13] Z. Wang webpage, "http://www.cns.nyu.edu/ zwang/," .