MEASURING THE ADDED HIGH FREQUENCY ENERGY IN COMPRESSED VIDEO

Athanasios Leontaris, Pamela C. Cosman

University of California, San Diego
La Jolla, CA 92093
{aleontar,pcosman}@code.ucsd.edu

Amy R. Reibman

AT&T Labs - Research
Florham Park, NJ 07932
amy@research.att.com

ABSTRACT

A major focus of video quality assessment research has been to quantify the amount of blocking, blurring, and ringing impairments. However, little attention has been paid to another impairment common in motion-compensated video compression systems: the addition of high frequency (HF) energy as motion compensation moves blocking artifacts off block boundaries. In this paper, we employ an energy-based approach to measure this motion-compensated edge artifact (MCEA) impairment, using both compressed bitstream information and decoded pixels. Experimental results show that we can accurately estimate the percentage of this energy in compressed video.

1. INTRODUCTION

Standardization bodies such as the Video Quality Experts Group [1] have been coordinating research efforts towards designing an efficient objective video quality metric. The goal is the automatic prediction of perceived image and video quality. Video quality metrics are used not only to assess the quality of reconstructed video, but also for fine-tuning and design of video coding systems. Peak Signal-to-Noise Ratio (PSNR) and Mean-Squared Error (MSE) have seen widespread use as video quality metrics due to their implementation simplicity and adequate performance. Unfortunately they do not take into account the perceptual characteristics of the Human Visual System (HVS). Incorporating HVS models into video quality metrics, as proposed in [2], is highly desirable as previous research [3] has shown that the widely-used PSNR metric cannot perform well in video sequences with significant luminance or texture masking.

No-reference metrics have access only to the reconstructed video sequence and its bitstream. These metrics are universally deployable, since they do not require access to the original sequence. Subjective tests done with humans are typically used as the ground truth to verify the results of an objective metric. It was shown in [4] that a combination of carefully crafted expectations satisfied by well-behaving metrics, in addition to a few small-scale subjective tests, can identify poorly behaving video quality metrics. These expectations included, among others, the ability to resolve the increase in blurriness and the decrease in similarity as the Quantization Parameter (QP) increases. Available metrics were unable to recognize that visual quality degrades as the distance from the last I-frame, *d*, increases.

Video quality is multidimensional. There are spatial and temporal dimensions as discussed in [5]. In this work we treat the spatial component of video quality. Visual quality spatial impairment is mainly constituted by three components: Blocking, blurring, and ringing. The estimation of those three components with the help of HVS models was the scope of [2]. All three can be encountered in both compressed still images and compressed video. Most research work on video quality assessment has concentrated on measuring blocking and blurring.

These impairment components are not completely orthogonal to each other. In [6] it was shown that the relative strength of one component can influence the perceptual contribution of another component. Research is still needed to characterize and quantify these interactions. One approach to quality assessment involves the use of HVS principles to determine a single value-index that characterizes the overall video quality [7, 8]. An alternative is to design metrics that assess a single impairment type, such that the impact of multiple impairments can be subsequently combined into a single quality value [9, 10].

In this work we adopt the second approach. The video quality metrics evaluation in [4] investigated the performance of state-of-the-art blocking and blurring metrics, and pointed to inadequacies of quality metrics when applied to motion-compensated video codecs. The source of these problems are motion-compensated edge artifacts (MCEA). The evaluated metrics were primarily developed for use with image codecs; thus they ignored this component of visual impairment. The MCEA is a side effect of the blocking impairment and motion-compensated prediction. Subjective tests in [4] showed that the perceptual effect of this unexplored visual impairment was significant.

The paper is organized as follows: Section 2 defines motion-compensated edge artifacts and discusses our motivation to design a metric to measure this type of impairment. The proposed metric is described thoroughly in Section 3. Experimental results are presented in Section 4 and the paper concludes in Section 5.

2. DEFINITION AND MOTIVATION

Video quality is a function of four types of impairments: blocking, blurring, ringing, and motion-compensated edge artifacts.

Blockiness arises from the vertical and horizontal edges along a regular blocking grid that result from the block-based processing in many image and video codecs. Coarse quantization yields more blockiness, while edge-attenuating filters reduce its perceptual effect. In this study we concentrate on the 8×8 DCT transform since it has seen widespread deployment in JPEG and MPEG.

Blurriness is caused by the removal of high-frequency content from the original image/video signal. Increased blurriness can be caused by coarser quantization, edge-attenuating filters, fractional-pixel motion compensation (MC) or overlapped block motion compensation (OBMC).

0-7803-9134-9/05/$20.00 ©2005 IEEE

Ringing artifacts, also known as the Gibbs phenomenon, are caused by the absence of high frequency terms from Fourier series due to coarse quantization. Perceived as ripples and overshoots near high contrast edges, they are most prevalent in wavelet coders.

Motion-compensated edge artifacts (MCEA) appear in video codecs that use block-based MC prediction. When coarse quantization is combined with MC prediction, blocking artifacts propagate from I-frames into subsequent frames and accumulate, causing structured HF noise that is no longer located at block boundaries. Fractional-pel MC and edge-attenuating filters can reduce this artifact. By definition, the MCEA involves HF noise within the blocks, while the blocking impairment involves HF noise along the block boundaries. The artifacts were called "false edges" in [11].

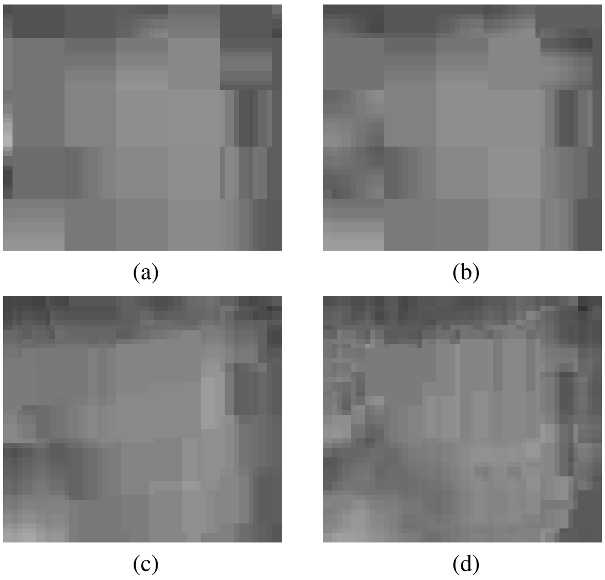


Fig. 1. Visual example of propagating motion-compensated edge artifacts in "foreman". (a) I-frame, (b) P-frame $d = 1$, (c) P-frame $d = 6$, (d) P-frame $d = 14$.

One example of MCEA is seen in Fig. 1. This 42×37 pixel segment, with its top left corner centered at the $(207, 155)$ pixel of frame 21 of "foreman" CIF, was encoded with QP set to 22. The same frame is encoded first as an I-frame, and then as a P-frame with $d = \{1, 6, 14\}$. In Fig. 1(a) the regular blocking grid is well perceived as it is a portion of an I-frame. For P-frame coding with $d = 1$ in Fig. 1(b) we observe that it looks similar to the previous case with minor spatial displacements in some block edges. For both Fig. 1(a) and 1(b) the spatial content within the block has low spatial frequency. As the distance d from the last I-frame increases, we observe significant changes in Fig. 1(c)-(d). Not only do block boundaries of the blocking grid dissipate, but new high frequency artifacts appear within the block boundaries, that are not part of the original image content.

The video quality metrics evaluation presented in [4] compared several blocking and blurring metrics. Among others, the blocking metrics were evaluated in their ability to order frames encoded with the same QP as I-frames, or as P-frames with increasing distance d from the last transmitted I-frame. Apart from a few full-reference metrics (MSE, PIQE-B [12] and SSIM [13]), none of the evaluated no-reference metrics resolved the difference successfully. Subjective testing showed that as d increases, the test subjects perceived an increase in "blockiness", which was a combination of both blockiness and MCEA. Most blocking metrics concentrate on the block boundaries, having been primarily designed with still images in mind, while humans are not as good at assessing if a particular artifact lies on a block boundary or not.

Traditional methods are not designed to measure these artifacts in P-frames. Pixel-based [14, 15] metrics require exact knowledge of artifact location, which is difficult to achieve due to the combination of fractional-pel motion compensation and variable-sized blocks. It would also be difficult to modify frequency-based blocking metrics [16, 17] to measure these artifacts. These methods rely on the periodicity of the blocking grid (see Fig. 1(a)).

3. MOTION-COMPENSATED EDGE ARTIFACT METRIC

Let M denote the measured DCT energy extracted from an 8×8 block in the decoded picture, C denote the energy calculated from the residual 8×8 block DCT coefficients transmitted in the compressed bitstream, and P denote the prediction 8×8 block energy. Both P and C can be computed exactly given the decoded pixels m and the transmitted coefficients c, since the predicted pixels $p = m - c$. We seek to design a no-reference metric that estimates the percentage of high frequency energy in M that is not part of the original image content. This added energy is a result of P not being an accurate estimate of the HF energy of the unknown source block, S, in the current frame. The encoder selected the prediction block because it was the best fit overall, but still its energy P may not accurately estimate the HF energy in S. The actual starting amount of extra HF energy is the HF energy in $P - S$: the energy of the non-quantized original residuals. Thus, the encoder compresses and transmits the residuals with available bits, resulting in quantized residuals with energy C. The HF energy in the bitstream, C, clearly, only *reduces* the HF error. Hence, the added HF error can be estimated as $(P - S) - C$. We now need to estimate the source block energy S.

Here, we estimate S using the weighted average energy E of the four blocks (aligned with the blocking grid) in the past frame that are used to form the prediction P. E is also the estimate of how much HF energy would have been in S, had S been sent using an I-frame. These four blocks in the past frame overlap the prediction block with energy P. The assumption here is that E is a good estimate due to *local stationarity*; i.e. the energy of regular grid blocks in a local neighborhood does not change significantly from one frame to the next. Thus, our estimate of the MCEA energy added by MC when encoding this frame can be written as $(P - E) - C$. We note that the energy estimate $(P - E) - C$ is not the same as the energy in the estimated signal $(p - e) - c$. We estimate its energy E but not the actual signal. We note the following assumption: the reconstructed residuals c are *uncorrelated* with the original signal s. Recursion is necessary to include the effects from previous frames. The final estimate of the frame MCEA energy is then normalized to incorporate texture masking.

Our approach involves the calculation of the MCEA energy on a block basis. A block-based approach addresses the occurence of skip blocks, where no DCT coefficients are transmitted at all, with the efficient use of recursion. For example, if the current block is a skip block, the MCEA energy is set to the one calculated for the co-located one in the previous frame. Similarly, for intra blocks we merely set it to zero as it is by definition. The detailed implementation follows.

Let the set of all 8×8-pixel blocks in a frame be T. The DCT coefficient in the compressed bitstream (transmitted prediction residuals) at location (i, j) of an 8×8 block $\tau \in T$ in frame n is $c_\tau^n(i, j)$. The DCT coefficient obtained from the 8×8 DCT transform of the reconstructed frame is similarly denoted (for the same spatial position) as $m_\tau^n(i, j)$. Here, the set of coefficients we consider, N, is the set of all AC DCT coefficients.

To estimate the source energy for a given block, τ, we let $\sigma(\tau)$ indicate the set of (up to) four blocks (aligned with the transform blocking grid) in frame $n-1$ that are used to predict block τ in frame n. The prediction of τ uses $w(\beta)$ percent of the block $\beta \in \sigma(\tau)$. Then, the energy estimate:

$$E_\tau^n = \sum_{\beta \in \sigma(\tau)} w(\beta) \sum_{(i,j) \in N} \left(m_\beta^{n-1}(i, j)\right)^2 \tag{1}$$

approximates the energy content of the source for the block τ in the current frame.

Now, we can compute the energy P_τ^n in the actual prediction block exactly, using the measured DCT coefficients in the reconstruction, $m_\tau^n(i, j)$, and the received coefficients, $c_\tau^n(i, j)$. Prior to adding the residual signal, c, the HF energy added by MC can be estimated as $B_\tau^n = P_\tau^n - E_\tau^n$ which we rewrite as:

$$B_\tau^n = \sum_{(i,j) \in N} \left(m_\tau^n(i, j) - c_\tau^n(i, j)\right)^2 - E_\tau^n \tag{2}$$

However, not all of the above energy ends up in the reconstructed frame. The transmitted DCT energy $C_\tau^n = \sum_{(i,j) \in N} \left(c_\tau^n(i, j)\right)^2$ serves only to improve the image quality and decrease the MCEA energy. The MCEA energy contribution for block τ in frame n can finally be estimated as:

$$H_\tau^n = B_\tau^n - C_\tau^n \tag{3}$$

Note that Eq. 3 can be negative, indicating that the transmitted energy C_τ^n was enough not only to counter the potential new MCEA energy B_τ^n but also to offset previously propagated MCEA energy. H_τ^n has to be added to MCEA propagated from previous frames. We thus obtain the recursive metric:

$$\mu_\tau^n = H_\tau^n + \sum_{\beta \in \sigma(\tau)} w(\beta) \mu_\beta^{n-1} \tag{4}$$

The second term on the right hand side is the propagated MCEA energy from the previous referenced blocks. We now define the measured energy content of the frame:

$$M^n = \sum_{\tau \in T} M_\tau^n = \sum_{\tau \in T} \sum_{(i,j) \in N} \left(m_\tau^n(i, j)\right)^2 \tag{5}$$

In those few cases that the transmitted energy C_τ^n is found to be greater than the maximally added HF energy B_τ^n and greater than the increase in local energy (measured energy M_τ^n minus the estimated energy E_τ^n), the calculated metric μ_τ^n for the block is set to be zero disregarding previously accumulated energy. Intuitively, if the transmitted DCT energy was enough to offset B_τ^n, and was again larger than the increase in local energy, we can speculate that it was enough to counterbalance all previously propagated MCEA energy (since the increase in energy can be solely attributed to the image content). The final metric can now be written as:

$$MCEA = \frac{\mu_\tau^n}{M^n} \tag{6}$$

which is an estimate of the percentage of DCT energy in the reconstructed video frame that is caused by MCEA.

4. RESULTS

We employed MEncoder H.263+ [18] to compress the sequences "foreman", "coastguard", "mother-daughter", and "mobile- calendar". We vary the QP from 2 (high quality) to 30 (low quality), and to minimize the impact of spatial content on the results, we compute the metric on the same frame 21 encoded using different distances from the last I-frame $d = 1, 6, 14$. Results are presented in Fig. 2(a)-(d). Let us now describe our expectations and discuss the metric's performance.

(a) For the same QP and filtering we expect that the energy of MCEA increases with the distance d from the most recent I-frame. We observe in Fig. 2 that the metric captures this for all sequences.

(b) For the same QP and d we expect that in-the-loop filtering decreases the energy of MCEA. Indeed, the metric is lower when filtering is used.

(c) For the same distance d and in the absence of filtering, we expect the curves to be monotonically increasing as the QP increases. This expectation is satisfied for the majority of sequences. The metric is highly irregular only for "mother-daughter". In fact, this sequence is so low motion and so easy to encode that it hardly has any perceivable MCEA, either perceptually or numerically. Thus, our metric correctly showed that the percentage of MCEA energy is extremely low.

(d) Subjective testing for the same QP and d showed that "foreman" and "coastguard" have the most visible MCEA. "Mobile" has significantly fewer MCEA, because they are masked by the abundance of HF image content. The MCEA in "mother" are almost invisible. Our metric correctly rank-ordered the sequences to match this subjective evaluation.

In addition to the expectations discussed above we sought a further reference for comparison. We designed a simple full-reference (FR) metric that calculates the energy difference on a block basis between the reconstructed video and the original sequence. The FR metric calculates the actual added HF energy due to MC. The correlation coefficients between our method and the FR metric, for the four sequences: "foreman", "coastguard", "mother- daughter", and "mobile- calendar", were obtained as: 0.9294, 0.9624, -0.5250, and 0.8638. The negative values for "mother" are explained in the discussion of expectation (c). It seems that our method has adequate correlation to FR, and hence estimates the energy of MCEA.

5. CONCLUSIONS

We discussed and defined a new component of visual impairment, which, while being ubiquitous in modern block-based video codecs, had not been investigated before. This impairment is termed motion-compensated edge artifact and is a direct consequence of motion-compensated prediction. High frequency energy that is not a part of the original image content is added and accumulated in consecutive P-frames. This energy is perceived as irregular artifacts within the original blocking grid. We presented a measurement framework based on calculating and estimating DCT energies in the current block and its local neighborhood in the previous frame. Experimental results proved both the accuracy of our metric and the efficiency of a measurement framework based on energy of DCT coefficients. Future work can include additional subjective testing, as well as investigating the design of metrics for the other impairments using this same framework.

6. REFERENCES

[1] VQEG, Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, http://www.vqeg.org/, Mar. 2000.

[2] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for image artifacts based on human visual sensitivity," in *Proc. IEEE ICASSP*, Apr. 1994, vol. 4, pp. 117–120.

[3] B. Girod, "What's wrong with mean-squared error?," in *Digital images and human vision*. MIT Press, 1993.

[4] A. Leontaris and A. R. Reibman, "Comparison of blocking and blurring metrics for video compression," in *Proc. IEEE ICASSP*, Mar. 2005, vol. 2, pp. 585–588.

[5] J. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video," in *Proc. CHI*, Apr. 2004.

[6] M. C. Q. Farias, S. K. Mitra, and J. M. Foley, "Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance," in *Proc. IEEE ICME*, 2003, vol. 1, pp. 529–532.

[7] C. J. Van den Branden and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of human video system," in *Proc. SPIE EI, vol. 2668*, 1996, pp. 451–461.

[8] M. A. Masry and S. S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *SP: Image Comm., Sp. Issue on obj. video qual. metrics*, vol. 19, pp. 133–146, Feb. 2004.

[9] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifacts in digital video," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 154–169, Jan. 2002.

[10] K. T. Tan and M. Ghanbari, "A multi-metric objective picture-quality measurement model for MPEG video," *IEEE Trans. CSVT*, vol. 10, no. 7, pp. 1208–1213, Oct. 2000.

[11] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, pp. 247–278, 1998.

[12] R. W. Chan and P. B. Goldsmith, "A psychovisually-based image quality evaluator for JPEG images," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, 2000, pp. 1541–1546.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Im. Proc.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

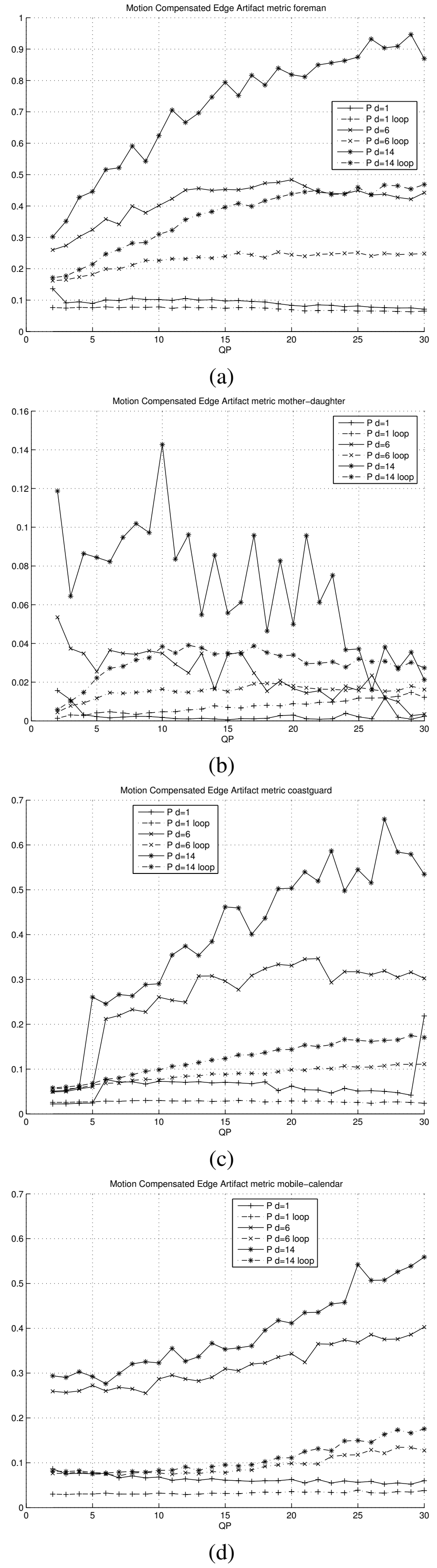
[14] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, no. 11, pp. 317–320, Nov. 1997.

[15] S. Suthaharan, "Perceptual quality metric for digital video coding," *IEE Electronics Letters*, vol. 39, no. 5, pp. 431–433, Mar. 2003.

[16] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE International Conference on Image Processing*, 2000, vol. 3, pp. 981–984.

[17] T. Vlachos, "Detection of blocking artifacts in compressed video," *IEE Electronics Letters*, vol. 36, no. 13, pp. 1106–1108, June 2000.

[18] MPlayer 1.0-pre4 software, http://www.mplayerhq.hu/.



(a)

(b)

(c)

(d)

Fig. 2. Experimental results are shown for the block-based metric. (a) "Foreman", (b) "Mother-Daughter", (c) "Coastguard", (d) "Mobile-Calendar".