BLIND ESTIMATION OF BLOCKING ARTIFACTS IN DIGITAL VIDEOS

Hugo Tadashi M. Kussaba and Mylène C.Q. Farias*

Department of Computer Science

University of Brasília (UnB), Brasília, DF, 70910-900, Brazil

ABSTRACT

In this paper, we present a no-reference video quality metric based on individual measurements of the blockiness artifact. The proposed metric is a modification of the work by Vlachos that measures the strength of blockiness artifacts in video sequences. Vlachos' algorithm estimates the blockiness signal strength by comparing the cross-correlation of pixels inside (intra) and outside (inter) the borders of the coding blocking structure of a frame. The proposed algorithm estimates the amount of spatial activity of each frame and uses this measure to modify the blockiness estimates according to its visibility. Contrary to what has been done by Vlachos, we used a Minkowski metric to combine the blockiness estimates for all frames of the video. The proposed algorithm was tested using a set of standard resolution videos compressed with an MPEG-2 enconder.

INTRODUCTION

In the past few years, considerable attention has been paid to the development of better video quality metrics that correlate well with the human perception of quality [1, 2]. Although many metrics have been proposed, most of them are very complex and require the original video for estimating the quality – Full Reference (FR) metrics. As a result, their use in real-time transmission applications is very difficult. For these applications, the solution is to use no-reference (NR) quality metrics, i.e., metrics that do not require the original (reference) to estimate quality.

Although human observers can usually assess the quality of a video without using the reference, designing a NR metric is a difficult task. Nevertheless, previous works have shown that it is possible to predict the overall annoyance of an impaired video using a combination of perceptual strengths of individual artifacts [3]. This means that we can estimate the quality of a degraded video by combining physical measurements (artifact metrics) of the most relevant artifacts. Among the currently available no-reference video quality metrics that use this approach, we can cite the works of Farias and Mitra [4] and Tan and Ghanbari [5].

Blockiness is one of the most relevant artifacts present in video applications. It is characterized by the visibility of the underlying block encoding structure and is often caused by a coarse quantization of the spatial frequency components during the encoding process [6]. A number of blockiness metrics have been proposed in the literature. Most of these metrics try to estimate blockiness spatially by detecting the edges in the

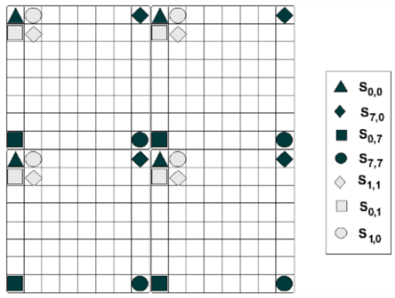

Fig. 1: Frame sampling structure for correlation-based blockiness metric in both horizontal and vertical directions.

frame corresponding to blockiness [7, 8].

In this work, we propose a no-reference blockiness metric for estimating the strength of the video impairments caused by digital compression. The proposed metric is a modification of the metric proposed by Vlachos [9], which uses a pixel-correlation approach. The algorithm proposed in this paper exploits the texture information present in the video frames to add a perceptual layer to the Vlachos' metric, in a way that it correlates better with the quality perception.

VLACHOS' BLOCKINESS METRIC

Vlachos' algorithm estimates the blockiness signal strength by comparing the cross-correlation of pixels inside (intra) and outside (inter) the borders of the coding blocking structure of a frame [9]. The algorithm considers that the size of the enconding blocks is b_s x b_s, with $b_s = 8$.

In Vlachos' work, the frame $Y(i,j)$ is partitioned into blocks and sampled to yield sub-images, given by:

$$s(m,n)=\{ Y(i, j): m = i \bmod b_s ,\ n = j \bmod b_s\}, \quad (1)$$

where (i, j) are frame pixel co-ordinates and $x \bmod y$ denotes congruence (remainder of integer division x/y).

The sub-image $s(m,n)$ contains the subset of pixels which are congruent with respect to block size. We can think of $s(m,n)$ as a sub-image obtained from sub-sampling the frame Y by pixels in both horizontal and vertical directions. Clearly, if before downsampling a shift is given to the frame Y, i.e., $Y_s = Y(i+m, j+n)$, different sub-images will be generated. This shift can be understood as a sampling phase. We represent a sub-image with sampling phase (m,n) by $s_{m,n}$.

*This work was supported in part by Conselho Nacional de Desenvolvimento Cientfico e Tecnológico - CNPq-Brazil and in part by a grant ProIC/DPP from the University of Brasília.

To estimate blockiness, seven sub-images with different sampling phases are considered. Figure 1 displays a zoom of this sampling structure where the different symbols represent a pixel of each different sub-image. The set composed of the pixels in sub-images $s_{0,0}$, $s_{0,7}$, $s_{7,0}$, and $s_{7,7}$ make out the set of inter-block pixels, while the set composed of pixels in $s_{0,0}$, $s_{0,1}$, $s_{1,0}$, and $s_{1,1}$ make out the set of intra-block pixels.

The correlation between a pair of images provides a measure of their similarity. To measure the correlation between two given images, *x* and *y*, we first calculate the correlation surface [10] using the following expression:

$$C_{x,y}=F^{-1}\left(\frac{F^{*}(x)\cdot F(y)}{\left|F^{*}(x)\cdot F(y)\right|}\right), \quad (2)$$

where *F* and *F** denote the forward and inverse two dimensional discrete Fourier transform, respectively, and * denotes the complex conjugate.

For identical images, the correlation surface has a unique peak, which is the two dimensional Dirac delta function. For non-identical images, which is usually the case, several peaks can be simultaneously present. The magnitude of the highest peak is used as a measure of correlation between *x* and *y* [10] :

$$p(x,y)=\max_{i,j}\{C_{x,y}(i,j)\}, \quad (3)$$

where (*i*, *j*) are the horizontal and vertical co-ordinates.

One problem with the above equation is that the periodic nature of the Fourier transform introduces sharp transitions at the borders [11]. So, before the maximum is taken, it is necessary to filter $C_{x,y}$ using a Hamming window to force the elements to a constant value around the borders.

To estimate the blockiness signal strength, we measure the correlation between the intra- and inter-block sub-images. In other words, we find the highest peaks of the phase correlation surfaces computed between the pairs of subimages. Considering the following subimages s_0 = s(0,0), s_1 = s(0,1), s_2 = s(1,0), s_3 = s(1,1), s_4 = s(0,1), s_5 = s(7,7), s_6 = s(0,7), s_7 = s(7,0), s_8 = (7,7), the blockiness measure is given by the ratio between a measure of intra-block similarity and a measure of inter-block similarity:

$$B=\frac{P_{intra}}{P_{inter}}, \quad (4)$$

where

$$P_{intra}=\sum_{i=1}^{3} p_{0,i} \quad \text{and} \quad P_{inter}=\sum_{i=6}^{8} p_{5,i}.$$

As blockiness is introduced, the values of become smaller and, consequently, the value of *B* increases. In Vlachos' original work, the blockiness measure for the set of all frames is obtained by averaging the measures over all frames:

$$\tilde{B}=\frac{1}{K}\sum_{k=0}^{K} B(k), \quad (5)$$

where the index *k* refers to the frame number and *K* is the total number of frames.

PROPOSED ALGORITHM

If placed in a region of high texture, the blockiness artifact may be perceptually masked. In other words, the annoyance caused by a blockiness artifact in a high texture region is smaller than the annoyance caused by the same artifact on a region of low texture. Therefore, we can use a texture measure to give perceptual weights to the blockiness measure obtained in each frame.

We quantify the amount of texture in a region by measuring the spatial variation of the pixel intensities in this region. Then, we classify a region with a high spatial variation (possible texture) as 'rough' and a region with low spatial variation as 'smooth'. A good way to estimate spatial variation of the pixel intensities in a single region is to use the standard deviation of the pixel intensities in this region. But, in order to assure that the blockiness of the frame does not interfere with the measure of the texture, we calculate the standard deviation of 8x8 pixels regions aligned with the encoding structure, what garantees that the block borders are not included.

The algorithm used to measure the *texture index* (T) follows:

1. Initialize the *texture index* of the *k*-th frame with 0, i.e., *T*(*k*)=0.
2. Calculate the standard deviation of each block of the *k*-th frame $(\sigma_i,\ 1\leq i\leq N)$.
3. Calculate the mean of the standard deviation of all *N* blocks of the *k*-th frame:

   $$\tilde{\sigma}(k)=\frac{1}{N}\sum_{i=0}^{N}\sigma_i.$$

4. Compare the standard deviation of each block $\sigma_i(k)$ with $\tilde{\sigma}(k)$. If the standard deviation of the block is greater than the mean standard deviation of the frame by a constant value α $(\alpha\geq 1)$, then the *texture index* is incremented by 1 (*T*(*k*)=*T*(*k*)+1). In our simulations we used α = 2.

Figures 2, 3, and 4 show representative frames of the three videos (*bike*, *football*, and *cargate*) used in simulations. The values of the *texture index* obtained for these frames are 1355, 1118 and 958, respectively. This result is in agreement with our perception of the amount of spatial activity in these frames. For example, the video *bike* has more regions with greater spatial differences than the video *football*, which has a uniform football field, and the video *cargate*, which has a 'smooth' road.

To perceptually improve Vlachos' blockiness metric, we use the *texture index* described above to model the effect of the video content masking the blocking artifact. The proposed algorithm gives a 'texture-aware' blockiness measure.

Fig. 2: Sample frame of 'bike' video

Fig. 3: Sample frame of 'cargate' video

Fig. 4: Sample frame of 'football' video

For each (*PB*) of the *k*-th frame of the video using the following equation:

$$PB(k) = \frac{B(k)}{S} \frac{N_{blocks}}{T(k)}, \quad (6)$$

where N_{blocks} is the total number of 8x8 blocks of the video and *S* is a scale factor. In the above equation, the factor is used to map the minimum value of the perceptual blockiness (*PB*) to the minimum value of Vlachos' blockiness metric (*B*). We also set the scale factor *S*=10 in order to make the range of the proposed metric approximately the same as the range of Vlachos' metric.

To combine the *perceptual blockiness* measurements for all *K* frames of the video, we use the weighted Minkowski metric of order 3:

$$\overline{PB} = \sqrt[3]{\sum_{k=1}^{K} (PB(k))^3} \cdot \quad (7)$$

SIMULATION AND RESULTS

In this section, we present the results of the proposed *perceptual blockiness metric*. We tested the algorithm on a set of 3 videos (*bike*, *cargate*, and *football*) with standard definition (SD), spatial resolution 720x480, and progressive format. The *bike* and *football* videos are in 4:2:0 format and the *cargate* video is in 4:2:2 format. Figures 2, 3, and 4 show representative frames of the videos used in this work.

To test the metric, we compressed the original test sequences with an *MPEG-2* encoder at 6 bitrates: 0.5, 1, 2, 4, 8, and 10 Mbits/s. Table 1 presents the *PSNR* values for the sequences coded at these bitrates. The *PSNR* values for each video were calculated by taking the average of the *PSNR* values computed for all frames of the video.

To compare Vlachos' original metric and the perceptual blockiness metric, we plot the outputs of both metrics for each frame of the test sequences bike and cargate. Figures 5, and 6 correspond to Vlachos' blockiness, while Figures 7 and 8 correspond to the perceptual blockiness metric. We can notice from these graphs that, as expected, the curves corresponding to higher bitrates have smaller values than the curves corresponding to smaller bitrates (more compression). It is also possible to notice that, in general, the blockiness values have a certain degree variation with the frame number and the same bitrate, with two or more peaks as the videos progresses. Nevertheless, the curves maintain the same correspondence among them.

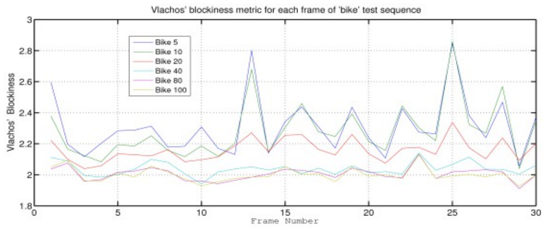

Fig. 5: Vlachos' Blockiness metric results for test sequence 'Bike' containing only blockiness.

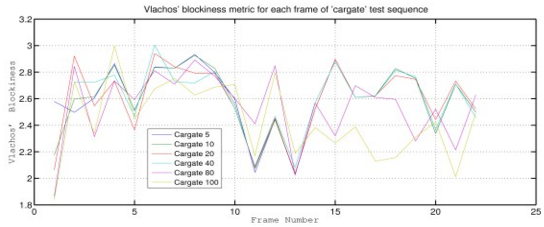

Fig. 6: Vlachos' Blockiness metric results for test sequence 'Cargate' containing only blockiness.

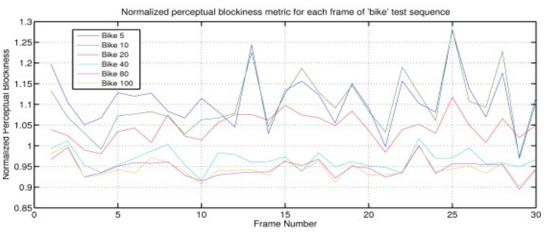

Fig. 7: Normalized Perceptual Blockiness metric results for test sequence 'Bike' containing only blockiness.

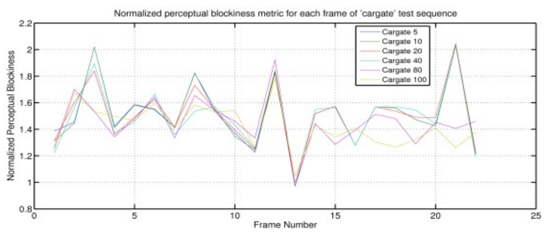

Fig. 8: Normalized Perceptual Blockiness metric results for test sequence 'Cargate' containing only blockiness.

In Table 1, we also present the results of both metrics for the whole video, for all test sequences. Again, as expected, the values of both metrics increase with the bitrate, indicating a decrease in blockiness strength. It is also possible to see that, for the videos 'cargate'and 'bike' the greater the texture index, is the faster is the decay of the metric with the bitrate (and consequently, the higher the quality of the video). In comparison, Vlachos' metric doesn't show this 'acceleration' in decay produced by the effect of texture. In other words, the proposed metric is more sensitivy to the effect of masking of the blocking artifact by texture.

Table 1: Values of the PSNR, Vlachos' metric (B), and Perceptual blockiness (PB) metric for the test sequences 'bike', 'cargate', and 'football' coded at bitrates 0.5, 1, 2, 4, 8 and 10 Mbps.

	Bitrate (Mbps)					
Bike	**0.5**	**1**	**2**	**4**	**8**	**10**
PSNR	25.87	26.44	29.3	33.52	37.55	38.76
B	2.31	2.28	2.16	2.04	2.01	2.00
PB	3.46	3.43	3.25	3	2.94	2.94
Car	**0.5**	**1**	**2**	**4**	**8**	**10**
PSNR	24.79	24.89	25.11	25.57	27.94	29.55
B	2.61	2.59	2.59	2.58	2.54	2.43
PB	4.34	4.33	4.31	4.29	4.13	4.04
Foot	**0.5**	**1**	**2**	**4**	**8**	**10**
PSNR	26.82	27.59	30.9	33.33	35.68	36.54
B	2.04	2.00	1.9	1.86	1.85	1.85
PB	3.18	3.17	3.18	3.21	3.25	3.28

CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a blind blockiness metric. The proposed metric is a modification of the work by Vlachos that estimates the blockiness by comparing the cross-correlation of pixels inside (intra) and outside (inter) the borders of the coding blocking structure. The proposed algorithm estimates the amount of spatial activity of each frame and uses this measure to modify the blockiness estimate according to its visibility. A Minkowski metric to combine the blockiness estimates for all frames of the video. The algorithm was tested using a set of SR videos compressed with an MPEG-2 enconder. As expected, the values of both metrics increase with the bitrate, indicating a decrease in blockiness strength.

The modifications added to Vlachos' blockiness metric are based solely on the effect of *artifact masking* produced by texture regions. These modifications can be easily applied to other blockiness metrics or to any artifact metric. Since the effect of masking is not negligible in most videos, this perception layer will probably increase the correlation of the quality metric with subjective quality scores. Nevertheless, further studies need to be completed in order to confirm this hypothesis. It is worth pointing out that 'pattern' masking is only one of the many aspects of the human visual system that can be exploited to perceptually improve the performance of video quality metrics.

REFERENCES

[1] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Processing*, vol. 78, no. 2, pp. 231–52, 1999.

[2] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment - Phase II," Tech. Rep., http://ftp.crc.ca/test/pub/crc/vqeg/, 2003.

[3] M.C.Q. Farias, J.M. Foley, and S.K. Mitra, "Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance," in *Proc. IEEE International Conference on Multimedia & Expo*, Baltimore, MD, USA, 2003, vol. 1, pp. 529–532.

[4] M.C.Q. Farias and S.K. Mitra, "No-reference video quality metric based on artifact measurements," in *Proc. IEEE Intl. Conf. on Image Processing*, 2005, pp. III: 141–144.

[5] K.T. Tan and M.Ghanbari, "A multi-metric objective picture-quality measurement model for MPEG video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 7, pp. 1208–13, 2000.

[6] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–78, 1998.

[7] 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, 1997.

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

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

[10] J.J. Pearson, D.C. Rines, S.Coldsman, and C.D. Kuglin, "Video rate image correlation processor," in *Proc. SPIE Conference on Application of Digital Image Processing*, San Diego, CA, 1977, vol. 119, pp. 197–205.

[11] G.A. Thomas, "Television motion measurement for dtv and other applications," Tech. Rep. 1987/11, BBC Res, Dept. Rep., 1987.