NO-REFERENCE VIDEO QUALITY METRIC BASED ON ARTIFACT MEASUREMENTS

Mylène C.Q. Farias and Sanjit K. Mitra

Department of Electrical and Computer Engineering
University of California , Santa Barbara, CA 93106 USA

ABSTRACT

In this paper we present a no-reference video quality metric based on individual measurements of three artifacts: blockiness, blurriness, and noisiness. The set of artifact metrics (physical strength measurements) was designed to be simple enough to be used in real-time applications. The metrics are tested using a proposed procedure that uses synthetic artifacts and subjective data obtained from previous experiments. The technique has the advantage of allowing us to test each metric on videos which contain only the desired artifact signal or a combination of artifact signals. Models for the overall annoyance based on a combination of the artifact metrics using both a Minkowski metric and a linear model are developed. Both models present a very good correlation with the data and show no statistical difference in their performances.

1. 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. As a result, their use in real-time transmission applications is very difficult. Although human observers can usually assess the quality of a video without using the reference, designing a *no-reference* metric is a difficult task and only a few of such video quality metrics have been proposed in the literature so far [3, 4].

In a previous work, we have found that it is possible to predict the overall annoyance of an impaired video using a combination of perceptual strengths of individual artifacts [5]. In this work, our goal is to investigate if a combination of physical strength measurements of individual artifact signals (*artifact metrics*) can be used to estimate the overall annoyance of impaired videos. The assumption here is that the strength of the artifact signal is correlated with the perceptual strength of the artifact. To this end, we developed a set of *no-reference* artifact metrics for blockiness, blurriness, and noisiness that are simple enough to be used in real-time applications. We, then, obtained models for overall annoyance using the Minkowski metric and the linear model. The proposed models focus on estimating the quality of standard definition videos compressed using MPEG-2 and its target application is broadcasting.

This work was supported in part by CAPES-Brazil, in part by a National Science Foundation Grant CCR-0105404, and in part by a University of California MICRO grant with matching support from Philips Research Laboratories.

2. ARTIFACT METRICS

The general approach for developing and testing artifact metrics consists of using video sequences with different levels of artifact strengths generated by compressing or by transmitting the original video at different bit-rates or conditions. In this work, we propose a different approach that provides a robust test and allows a better comparison among different metrics. To be able to control the type, proportion, and strength of the artifacts, we tested the artifact metrics using synthetically generated artifacts and subjective data gathered from specifically designed psychophysical experiments.

In this work, we used the data gathered from a previous experiment where blocky, blurry, and noisy artifacts were inserted into specific regions of the videos for a short time interval (defect regions). We asked subjects to detect them and rate their annoyance. A total of six originals were used - 'Bus', 'Calendar', 'Cheerleader', 'Flower', and 'Hockey' [2]. Three defect regions were used for each original to prevent the test subjects from learning the locations of the defects. We varied the artifact strength by scaling the pixel-by-pixel difference between the corrupted and the original videos [6]. Table 1 shows the total squared error (TSE) of the test sequences corresponding to the original video 'Bus' [1]. In this section we present the blockiness, blurriness, and noisiness metrics which performed better in our simulation tests.

2.1. Blockiness Metric

The proposed blockiness metric is a modification of the metric proposed by Vlachos [7]. 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. In his work, the frame $Y(i,j)$ is partitioned into $b_s \times b_s$ blocks and simultaneously sampled in vertical and horizontal directions. This sampling structure assumes that all visible blockiness artifacts have a visible corner. Nevertheless, frequently, only one of the borders of the blocking structure is visible. To reflect this, instead of down-sampling the frame simultaneously, we splitted the process into two separate parts. This modification improves the performance of the algorithm, with a small increase in complexity.

First, we down-sampled the frame only in the vertical direction, obtaining the vertical sub-sampled image s_v:

$$s_v(m,n) = \{Y(i,j) : m = i, n = j \bmod b_s\}. \qquad (1)$$

where (i,j) are the horizontal and vertical co-ordinates. Then, we down-sampled the frame in the horizontal direction, obtaining the horizontal sub-sampled image s_h:

$$s_h(m,n) = \{Y(i,j) : m = i \bmod b_s, n = j\}. \qquad (2)$$

[1] For lack of space, we will only show the results corresponding to the test video 'Bus'. The results obtained for other videos were similar.

Table 1. Total Square Error values for test sequences containing blocky, blurry, and noisy artifacts.

defect region	artifact strength	TSE		
		blocky	blurry	noisy
Top	1	109.09	40.18	157.97
Top	2	197.82	243.44	427.99
Top	3	396.44	621.94	1328.56
Top	4	670.46	1370.61	2622.14
Top	5	996.53	2382.24	4186.98
Top	6	1415.09	3626.78	6093.68
Middle	1	423.19	527.46	675.01
Middle	2	801.39	3283.43	1803.30
Middle	3	1563.37	8454.29	5474.06
Middle	4	2558.25	18919.50	10529.10
Middle	5	3771.61	33360.10	16317.20
Middle	6	5227.33	51554.00	22467.90
Bottom	1	436.23	241.95	227.86
Bottom	2	824.24	1499.86	614.08
Bottom	3	1651.10	3875.24	1899.68
Bottom	4	2752.66	8701.84	3726.34
Bottom	5	4092.39	15392.20	5899.88
Bottom	6	5715.65	23852.10	8397.98

The size of the block, b_s can be adjusted according to the application. In this work, $b_s = 8$ because we are targeting MPEG-2 codecs.

Figures 1 (a) and (b) display the modified sampling structures for the horizontal (s_h) and vertical (s_v) directions, respectively, for a 24×24 area of the frame. The dark symbols inside the grid correspond to pixels in the resulting sampled sub-images. Different symbols correspond to different sub-images. The set of inter-block pixels in the vertical direction corresponds to the sub-images s_7 and s_0 (Figure 1(b)), while the set of inter-block pixels in the horizontal direction corresponds to the sub-images s_7 and s_1 (Figure 1(a)). The set of intra-block pixels corresponds to the sub-images s_0 and s_1 for the vertical direction, and s_1 and s_3 for the horizontal direction.

The correlation between two given images, s_m and s_n, is given by the following expression:

$$C_{m,n}(i,j) = F^{-1}\left(\frac{F^*(s_m(i,j)) \cdot F(s_n(i,j))}{|F^*(s_m(i,j))F(s_n(i,j))|}\right), \quad (3)$$

where F and F^{-1} denote the forward and inverse two dimensional discrete Fourier transform, respectively, and * denotes the complex conjugate. The magnitude of the highest peak was used as a measure of correlation between s_m and s_n. Before the maximum was taken, the array elements were filtered using a Hamming window to force the elements to a constant value around the borders.

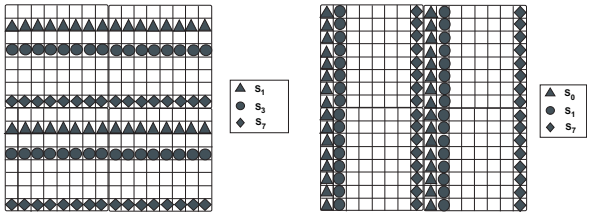
To estimate the blockiness signal strength, we measured the correlation between the intra- and inter-block sub-images in both directions. The vertical intra-block correlation is given by:

$$PV_{intra} = \max_{i,j}\{C_{7,0}(i,j)\}, \quad (4)$$

while the vertical inter-block correlation is given:

$$PV_{inter} = \max_{i,j}\{C_{0,1}(i,j)\}. \quad (5)$$

The horizontal correlations, PH_{inter} and PH_{intra}, are obtained in a similar way. The blockiness measure for the frame is given by the following expression:



(a) (b)

Fig. 1. Frame sampling structure for correlation-based blockiness metric: (a) horizontal and (b) vertical directions.

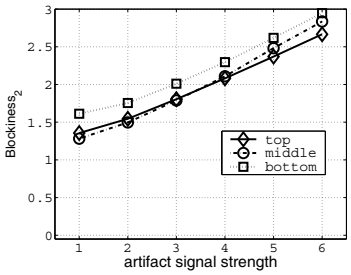


Fig. 2. Blockiness metric results for test sequence 'Bus' containing only blockiness.

$$\text{blockiness} = 2.0 - \left[\frac{PV_{intra}}{PV_{inter}} + \frac{PH_{intra}}{PH_{inter}},\right] \quad (6)$$

For frames with no blockiness, the value of PV_{intra} is close to PV_{inter} and PH_{intra} is close to PH_{inter}. As blockiness is introduced, the values of PV_{inter} and PH_{inter} become smaller and, consequently, the value of the blockiness metric increases. The blockiness signal measurement for the set of all frames was obtained by averaging the measures over all frames.

Figure 2 displays the results after applying this metric on test sequences containing only synthetic blocky artifact signals for the original 'Bus'. The *x*-axis of the graph correspond to the six blockiness signal strengths shown in the third column of Table 1 and represented, for simplification, by the integers 1-6, where 1 refers to the smallest strength and 6 to the largest one. The *y*-axis corresponds to the blockiness signal measurements. The three curves correspond to the three defect regions of the video ('top', 'middle', and 'bottom'). Similar results have been found for the other original videos. As can be seen, the blockiness metric increases as the blockiness signal strength increases. Simulation results showed that the metric was robust, had a good performance, and was not very sensitive to content effect. Overall, the metric performed better (the correlation with the subjective data was $r = 0.860$) than other metrics which estimate blockiness in the spatial [8] and frequency [4] domain ($r \approx 0.40$).

2.2. Blurriness Metric

Most of the existing blurring metrics are based on the idea that blur makes the edges larger or less sharp [9, 10, 11]. In this work, we implemented a *no-reference* blur (blurriness signal) metric which also makes use of this very simple idea. The algorithm measures blurriness by measuring the width of the edges in the frame. The first step consists of finding strong edges using the Canny edge detector algorithm. The output of the Canny algorithm gives the

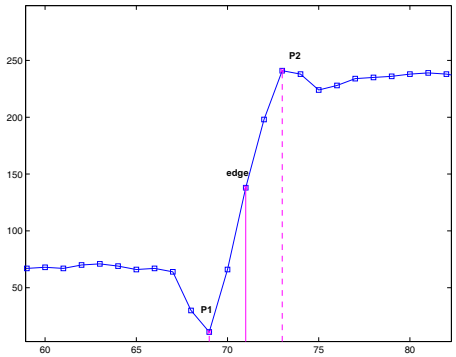


Fig. 3. The width of the edge is used as a measure of the blurriness signal strength. P_1 is the first local extreme and P_2 is the second one.

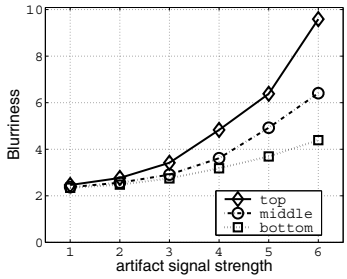


Fig. 4. Blurriness metric results for test sequences 'Bus' containing only blurriness for videos.

magnitude of the edge pixels, $M(i,j)$, and their orientation, $O(i,j)$. We selected only the strong edges of the frame ($M(i,j) > 25$).

The width of an edge is defined as the distance between the two local extremes, P_1 and P_2, on each side of the edge, as shown in Figure 3. If the edge is horizontal, P_1 will be located above the edge pixel, while P_2 will be below it. If the edge is vertical, P_1 will be located to the left of the edge pixel, while P_2 will be to the right of it. The width of the edge, $width(i,j)$, at position (i,j) is given by the difference between the two extremes $P_1(i,j)$ and $P_2(i,j)$. The blurriness signal strength measure for a frame was obtained by averaging widths over all strong edges of this frame. So, given that a frame Y has L strong edges pixels, the blurriness signal strength measure for this frame is given by:

$$\text{blurriness} = \frac{1}{L}\sum_{i=0}^{N}\sum_{j=0}^{M} width(i,j). \tag{7}$$

The blurriness signal strength measure for the whole video was obtained by averaging the measurements for all frames.

Figure 4 displays the results of applying this metric on the test sequence containing only blurriness for the original 'Bus'. The *x*-axis of the graphs corresponds to the six blurriness signal strengths (see the fourth column of Table 1) and the *y*-axis corresponds to blurriness metric. Each curve corresponds to one different region of the video. As can be seen from this figure, the blurriness metric increases as the artifact signal strength increases. This metric, although very simple, was very robust and insensitive to contents.

2.3. Noisiness Metric

Most existing noisiness metrics are designed for still images and, frequently, they require some information about the original [12].

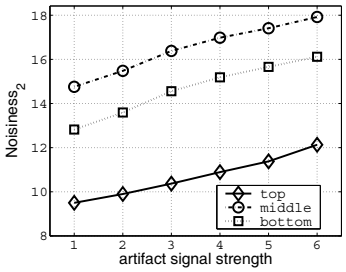


Fig. 5. Noisiness metric results for test sequence 'Bus' containing only noisiness.

Blindly estimating noise in an image or video is still a difficult problem. The proposed noisiness metric is based on the work by Lee [13] that uses the well known fact that the noise variance of an image can be estimated by the local variance of a flat area. To reduce the content effect we used a cascade of 1-D filters as a pre-processing stage [12].

After the preprocessing, we sub-divided the image in 8×8 blocks and calculated the local variance of each of the 9 overlapping 3×3 sub-blocks inside each block. An estimation $\widehat{\sigma}_l^2$ of the *l*-th block noise variance was calculated by averaging the 4 smallest sub-block variances. Then, we calculated the histogram $h(\widehat{\sigma})$ of the block variances. An initial global estimate for the frame noise variance was obtained by calculating the mean squared value of the histogram [13]:

$$s_1^2 = \frac{\sum_{k=0}^{k_{max}} k^2 h(k)}{\sum_{k=0}^{k_{max}} h(k)}. \tag{8}$$

Since the pre-processing stage does not do a perfect job in eliminating the content effect, Eq. (8) usually overestimates the variance. This problem was reduced by implementing a fade-out of the histogram using a simple cutoff function with threshold $\beta_1 = 1.5 \cdot s_1$.

An improved value of the mean squared s^2 was computed iteratively using the following expression:

$$s_{l+1}^2 = \frac{\sum_{\sigma=0}^{\sigma_{max}} \sigma^2 g_l(\sigma) h(\sigma)}{\sum_{\sigma=0}^{\sigma_{max}} g_l(\sigma) h(\sigma)}, \tag{9}$$

where σ_{max} is the maximum value for σ obtained while calculating the histogram. The initial value s_1 was taken from Eq. (8) and the mean squared value was refined successively. After three to five iterations, convergence was achieved. The final estimate of the frame noise variance and, therefore, of the noisiness signal strength, is given by $s_{l_{max}}^2$, where l_{max} indicates the total number of iterations used. The noisiness signal measure for the set of all frames was obtained by averaging the measures for each frame.

Figure 5 shows the results of applying this algorithm to the test sequence with only noisy artifacts for the original 'Bus'. The *x*-axis of the graphs correspond to the six noisiness signal strengths (see fifth column of Table 1) and the *y*-axis correspond to noisiness metric. As can be noticed, the noisiness metric increases as the signal strength of the noisiness increases. Notice that there is a considerable difference between the three curves corresponding to different regions of the video. This reflects the fact that the outputs of noisiness metrics are very influenced by the video content. Nevertheless, the correlation of the noisiness metric with the perceptual artifact strengths given by the subjects was good ($r = 0.74$) and the metric performed well for our purposes.

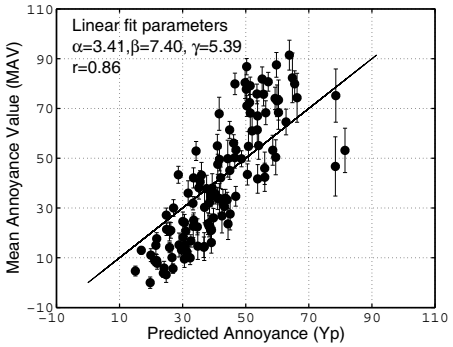


Fig. 6. Observed MAV versus predicted annoyance for the set of all test sequences for the linear model.

3. OVERALL ANNOYANCE ESTIMATION

If a video is affected by one or more types of artifacts, the total annoyance can be estimated from the individual artifact perceptual strengths using a combination rule. In this section, we want to investigate if the same type of model can be used to estimate overall annoyance by using, instead, individual artifact signal physical strength measurements (artifact metrics). So, after choosing the best artifact metrics, our goal is to obtain a model for overall annoyance using a combination rule of these metrics. We propose a *no-reference* annoyance model that uses the blockiness, blurriness, and noisiness metrics using the weighted Minkowski metric:

$$Y_p = (\alpha \cdot \text{Blockiness}^p + \beta \cdot \text{Blurriness}^p + \gamma \cdot \text{Noisiness}^p)^{1/p} \tag{10}$$

If $p = 1$, it becomes a linear model.

Then, we performed a nonlinear least-squares data fitting using both Minkowski metric and linear model to the mean annoyance values obtained from a second psychophysical experiment [5] and the output of the artifact metrics. This second psychophysical experiment independently measured the strength and annoyance of blocky, blurry, and noisy artifacts when presented alone or in combination. The data set consisted of 120 test sequences - 4 originals ('Bus', 'Cheerleader', 'Football', and 'Hockey') $\times$ 30 combinations of blocky, blurry, and noisy artifacts.

Both models produced a good correlation with the data (r = 0.86). Although the linear model is a simpler and more restrictive model, we found that there was no significant statistical difference, in performance, between the linear and the more generic Minkowski models. Figure 6 shows a plot of the measured versus predicted annoyance using linear metric. The fit for the Minkowski metric returned an exponent equal to 0.66 and scaling coefficients equal to 0.91, 3.40, and 2.51, corresponding to blockiness, blurriness, and noisiness, while for the linear model the scaling coefficients were equal to 3.41, 7.40, and 5.39, corresponding to blockiness, blurriness, and noisiness. The annoyance models using the artifact metrics had similar parameters to the ones found for the annoyance models obtained using the artifact perceptual strengths [5]. Further improvements in the models can be obtained by adding simple human visual systems to the metrics.

4. CONCLUSIONS

The goal of this paper was to test the possibility of predicting the overall annoyance of videos impaired with combinations of blockiness, blurriness, and noisiness metrics. With this purpose we designed a set of artifact metrics for blockiness, blurriness, and noisiness. To evaluate the performance of each artifact metric, we tested its ability to detect and estimate the artifact signal strength for test sequences containing only the artifact being measured. Finally, we obtained a model for overall annoyance based on a combination of the artifact metrics using both a Minkowski metric and a linear model. Both models presented a very good correlation with the data and no statistical difference in performance.

5. REFERENCES

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

[2] Video Quality Experts Group, "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] J. Caviedes and J. Jung, "No-reference metric for a video quality control loop," in *Proc. Int. Conf. on Information Systems, Analysis and Synthesis*, 2001, vol. 13.

[4] 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.

[5] 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.

[6] M.C.Q. Farias, J.M. Foley, and S.K. Mitra, "Detectability and annoyance of synthetic blockiness, blurriness, noisiness, and ringing in video sequences," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, 2005, vol. ii, pp. 553–556.

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

[8] 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.

[9] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.

[10] J. Lu, "Image analysis for video artifact estimation and measurement," in *Proc. SPIE Machine Vision Applications in Industrical Inspection IX*, San Jose, CA, USA, 2001, vol. 4301, pp. 166–174.

[11] E-P. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, and L. Jinag, "No-reference JPEG2000," in *Proc. IEEE International Conference on Multimedia and Expo*, Baltimore, USA, 2003, vol. 1, pp. 545–548.

[12] S. I. Olsen, "Estimation of noise in images: an evaluation," *CVGIP-Graphical Models & Image Processing*, vol. 55, no. 4, pp. 319–23, 1993.

[13] J.S. Lee and K. Hoppel, "Noise modeling and estimation of remotely-sensed images," in *Proc. International Geoscience and Remote Sensing*, Vancouver, Canada, 1989, vol. 2, pp. 1005–1008.