

# A HYBRID METRIC FOR DIGITAL VIDEO QUALITY ASSESSMENT

Mylène C.Q. Farias<sup>1</sup>, Marcelo M. Carvalho<sup>2</sup>, Hugo T.M. Kussaba<sup>1</sup>, and Bruno H.A. Noronha<sup>1</sup>

<sup>1</sup> Department of Computer Science  
<sup>2</sup> Department of Electrical Engineering  
University of Brasília (UnB),  
Brasília, DF, 70910-900, Brazil  
{mylene, carvalho}@ieee.org

## ABSTRACT

In this paper, we present a hybrid no-reference video quality metric. The proposed metric blindly estimates the quality of videos degraded by compression and transmission artifacts. The metric is composed by two no-reference artifact metrics that estimate the strength of blockiness and blurriness artifacts. A combination model is used to add the packet loss rate information to the quality estimate and eliminate the disturbance in the artifact metric values caused by the packet losses.

**Index Terms**— video quality metrics, artifacts, quality assessment, no-reference quality metrics, packet-loss, quality of service.

## 1. INTRODUCTION

Digital video communication has evolved into an important field in the past few years. There have been significant advances in compression and transmission techniques, which have made possible to deliver high quality video to the end user. In particular, the advent of new technologies has allowed the creation of many new telecommunication services (e.g., direct broadcast satellite, digital television, high definition TV, Internet video). In these services, the level of acceptability and popularity of a given multimedia application is clearly related to the reliability of the service and the quality of the content provided.

In this context, the term quality of experience (QoE) describes the quality of the multimedia service provided to the end user. Although there has been some debate regarding the actual meaning of this term, it is generally agreed that QoE encompasses different aspects of the user experience, such as video and audio quality, user expectation, display type, and viewing conditions. In this work, we are interested in estimating video quality according to user perception.

---

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil, in part by a grant by Fundação de Empreendimentos Científicos e Tecnológicos (Finatec) Brazil, and in part by a grant from DPP - University of Brasília (UnB).

The most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects (subjective metrics) [1]. Unfortunately, these experiments are expensive, time-consuming and hard to incorporate into a design process or an automatic quality of service control. Therefore, the ability to measure video quality accurately and efficiently, without using human observers, is highly desirable in practical applications. With this in mind, fast algorithms that give a physical measure (objective metrics) of the video quality are needed to obtain an estimate of the quality of a video when being transmitted, received or displayed.

As far as quality metrics are concerned, the networking community has been using simple metrics to quantify the quality of service (QoS) delivered to a given application, such as bit error rate (BER) or packet loss rate (PLR). Likewise, within the signal processing community, quality measurements have been largely limited to a few objective measures, such as peak signal-to-noise ratio (PSNR) and total squared error (TSE). Although these metrics are relevant for data links and generic signals in which every bit is considered equally important within the bitstream, they are not considered good estimates of the user's opinion about the received multimedia content [2, 3]. As a result, there is an ongoing effort to develop video quality metrics that are able to accurately detect impairments and estimate their annoyance as perceived by human viewers.

To date, most of the quality metrics proposed in the literature are Full Reference (FR) metrics [3], i.e., metrics that use the original to compute an estimate of the quality. FR metrics have limited applications and cannot be used in most real-time video transmission applications, like for example, broadcasting and video streaming. In such cases, the initial undistorted signal (reference) is not available or not accessible at the receiver side and, therefore, requiring the reference video or even a small portion of it becomes a serious impediment in real-time video transmission applications. To measure video quality in such applications, it is essential to use a no-reference (NR) or a reduced reference (RR) video

quality metric, i.e., a metrics that blindly estimates the quality of the video using no information (nr) or limited information about the original [4]. Most NR metrics have limited performance because estimating the quality or degradation without the original is a difficult task. In fact, in the Final Report of the Video Quality Experts Group (VQEG) Multimedia Phase I, it was found that the correlations for the submitted FR, RR, and NR metrics were around 80%, 78%, and 56%, respectively, corroborating the case that, to date, NR metrics still have poor performance [5].

In this paper, we propose a hybrid quality metric that consists of a combination of a QoS metric and an NR objective quality metric: the QoS metric takes into account packet loss rates, while the NR metric consists of two no-reference artifact metrics. The main advantage of this approach is the fact that it gives an estimate of quality without requiring the reference, while, at the same time, it uses additional, network-related information in order to leverage the NR metric performance. To assess the effectiveness of our hybrid metric, we evaluate the quality of H.264/AVC digital video transmissions subjected to packet loss patterns typical of the Internet backbone.

## 2. ARTIFACT METRICS

In this work, we focus on three of the most common artifacts present in digital videos: blockiness, blurriness, and packet loss. *Blockiness* is a type of artifact characterized by a block pattern visible in the picture. It is due to the independent quantization of individual blocks in block-based DCT coding schemes (usually,  $8 \times 8$  pixels in size), leading to discontinuities at the boundaries of adjacent blocks. The blocking effect is often the most visible artifact in a compressed video, given its periodicity and the extent of the pattern. Modern codecs, like the H.264, use a deblocking filter to reduce the annoyance caused by this artifact. *Blurriness* is characterized by a loss of spatial detail and a reduction of edge sharpness. In the compression stage, blurriness is introduced by the suppression of the high-frequency coefficients due to coarse quantization.

In video transmission over IP networks, video packets typically traverse a number of links to get to its destination. Packet losses may occur due to buffer overflow at network routers (caused by network congestion) or signal transmission/reception errors at the physical layer. Typical impairments caused by these errors are *packet-loss*, jitter, and delays. Among these, packet-loss is probably the most annoying impairment. As the name suggests, packet-loss impairments are caused by a complete loss of the packet being transmitted, as a consequence of transmission errors. As a consequence, parts (blocks) of the video are missing for several frames.

Figures 1(a) and 1(b) depict two sample frames of videos affected by packet-loss and a combination of blockiness and blurriness, respectively. The severity of the impairments can vary tremendously depending on the bitrate and network con-

ditions. The strength of blockiness and blurriness can be estimated using specific artifact metrics, while the strength of packet loss artifacts can be roughly estimated measuring the packet loss rate for the video at the receiver (QoS parameter). In this section, we present the two no-reference artifact metrics used to estimate *blockiness* and *blurriness*.



**Fig. 1.** Sample video frames containing medium and severe intensity packet-loss impairments.

### 2.1. 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 [6]. The algorithm considers that the size of the encoding blocks is  $b_s \times b_s$ , with  $b_s = 8$ . The frame  $Y(i, j)$  is partitioned into blocks and sampled to yield sub-images, given by:

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

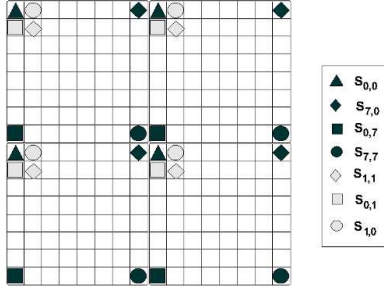
where  $(i, j)$  are frame pixel coordinates and  $x \bmod y$  denotes congruence (remainder of integer division  $x/y$ ).

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  $b_s$  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  $s_{m,n}$ .

To estimate blockiness, seven sub-images with different sampling phases are considered. Figure 2 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 [7] using the following expression:

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



**Fig. 2.** Frame sampling structure for correlation-based blockiness metric in both horizontal and vertical directions.

where  $F$  and  $F^{-1}$  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  $y$  [7] :

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

where  $(i, j)$  are the horizontal and vertical coordinates. One problem with this equation is that the periodic nature of the Fourier transform introduces sharp transitions at the borders [8]. 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  $s_0 = s(0, 0)$ ,  $s_1 = s(0, 1)$ ,  $s_2 = s(1, 0)$ ,  $s_3 = s(1, 1)$ ,  $s_4 = s(0, 1)$ ,  $s_5 = s(7, 7)$ ,  $s_6 = s(0, 7)$ ,  $s_7 = s(7, 0)$ ,  $s_8 = s(0, 0)$ , the blockiness measure is given by the ratio between a measure of intra-block similarity and a measure of inter-block similarity:

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

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

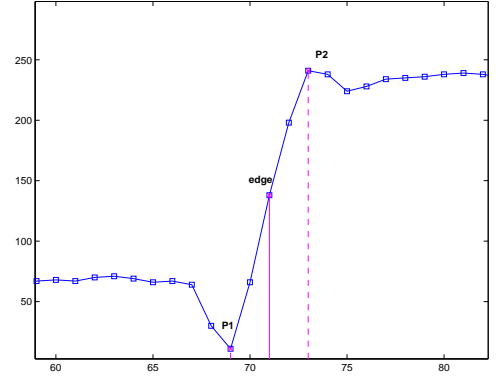
The more blockiness is introduced, the values of  $P_{inter}$  become smaller and, consequently, the value of Block increases. The blockiness measure for the set of all frames is obtained by taking the median of the measures over all frames.

## 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]. 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 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.



**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.

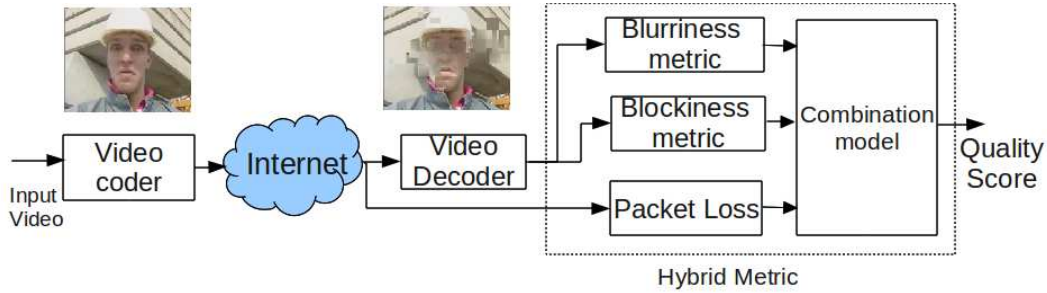
If the edge is horizontal,  $P_1$  will be located above the edge pixel, while  $P_2$  will be below it. If the edge is vertical,  $P_1$  will be located to the left of the edge pixel, while  $P_2$  will be to the right of it. The width of the edge,  $width(i, j)$ , at position  $(i, j)$  is given by 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:

$$\text{Blur} = \frac{1}{L} \sum_{i=0}^N \sum_{j=0}^M \text{width}(i, j). \quad (5)$$

The blurriness signal strength measure for the whole video is obtained by taking the median of the measures over all frames.

## 3. THE HYBRID QUALITY METRIC

In order to obtain a hybrid quality metric, we investigate the performance of each individual quality metric across a set of typical video sequences subject to different bitrates and packet loss levels. Once their individual performance is assessed, we propose a final combination model for the hybrid quality metric. Figure 4 summarizes the overall idea of the proposed hybrid metric.



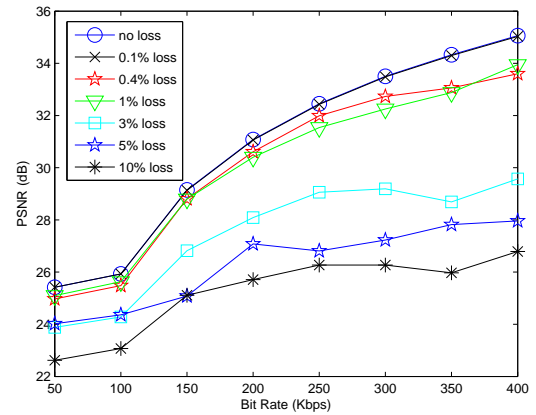
**Fig. 4.** Block diagram of the proposed hybrid quality metric.

For our study, we used publicly-available videos in CIF format ( $352 \times 288$  pixels), YUV 4:2:0 color format, with 300 frames. The videos we used were ‘foreman’, ‘mother’, ‘mobile’, ‘news’, and ‘paris’, all compressed with target bitrates of 50k, 100k, 150k, 200k, 250k, 300k, 350k, and 400k bps. In order to simulate packet losses in a given bitstream, we used the *transmitter simulator* [11], a software that simulates the transmission of H.264/AVC bitstreams over error-prone channels. For simulation of packet losses, the *transmitter simulator* makes use of error pattern files that are based on actual experiments carried out on the Internet backbone. The error pattern files correspond to packet loss rates (PLR) of 0.1%, 0.4%, 1%, 3%, 5%, and 10%, respectively. For analysis, we considered H.264/AVC bitstreams that were packetized according to the Real-Time Transfer Protocol (RTP). In simulations, all packets were treated equally regarding their susceptibility to loss (i.e., we did not focus on specific types of packets, such as those carrying intra-coded slices, for example).

To illustrate the quality range of the videos used in this work, Figures 5 and 6 show the Peak Signal to Noise Ratio (PSNR) values for different bitrates and PLR values for the videos foreman and paris. Observe that, for PLR values less than or equal to 1%, the PSNR values increase with the target bitrate. But, for PLR values greater than 1%, the PSNR values do not necessarily increase with the target bitrate.

Figures 7 and 8 depict the blockiness metric output values for the videos ‘foreman’ and ‘paris’, respectively, under different target bitrates, PSNR, and PLR values. Once again, notice that for PLR values equal or less than 1%, blockiness strength values decrease with the target bitrate (and PLR). But, for PLR values greater than 1%, blockiness strength values are disturbed by the packet losses and do not have a ‘reliable’ behaviour. Figures 9 and 10 depict the blurriness metric output values for the videos ‘foreman’ and ‘paris’, respectively, under different target bit rate, PSNR, and PLR values. We can notice from these graphs that the behaviour of the blurriness metric is more robust against the influence of PLR than the blockiness metrics.

Figure 11 depicts the relationship between packet loss rate (PLR) and PSNR values for the ‘foreman’ video sequence. In



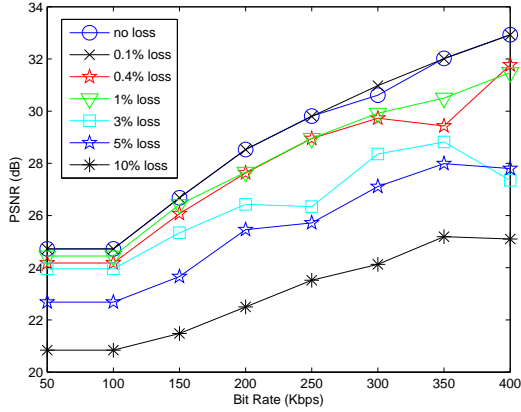
**Fig. 5.** PSNR values for ‘foreman’ video under different bit rates and packet loss rates.

this graph, each point corresponds to a different bitrate. As we can clearly observe, a zero PLR (or very low PLR value such as 0.1%, 0.4%, and 1.0%) does not necessarily mean a high-quality video, since it also depends on the coding scheme, as expressed by the wide range of PSNR values observed for these cases. On the other hand, as the PLR increases (3%, 5%, and 10%, respectively), not only the PSNR values decrease across all bitstreams, but their variability also decreases, indicating that PLR becomes a more consistent quality measure within this range, in spite of the coding scheme (similar behavior is also observed with the other video sequences). Therefore, it is exactly where the blockiness and blurriness metrics present their lower performance that the PLR becomes a more consistent measure of overall video quality. Based on such observations, we propose the hybrid quality metric  $Q$ , which is given by

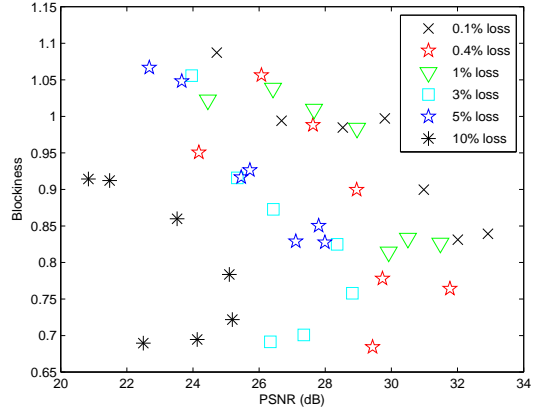
$$Q = (1 - \beta) f_1(\text{Blur}, \text{Block}) + \beta f_2(\text{PLR}), \quad (6)$$

where  $\beta$  is a weighting factor, and  $f_1(\text{Blur}, \text{Block})$  and  $f_2(\text{PLR})$  are quality estimators based on the blockiness, blurriness, and PLR metrics, respectively, given by

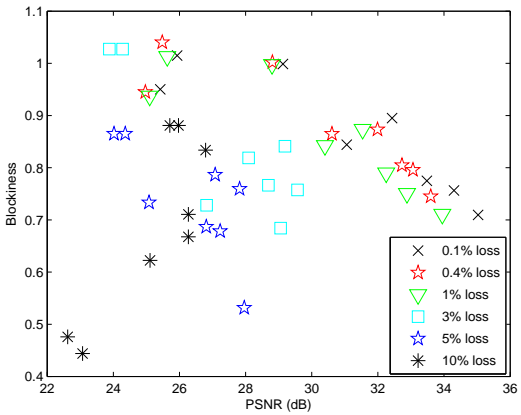
$$f_1(\text{Blur}, \text{Block}) = -14.7 \cdot \text{Block} - 1.1 \cdot \text{Blur} + 42.2 \quad (7)$$



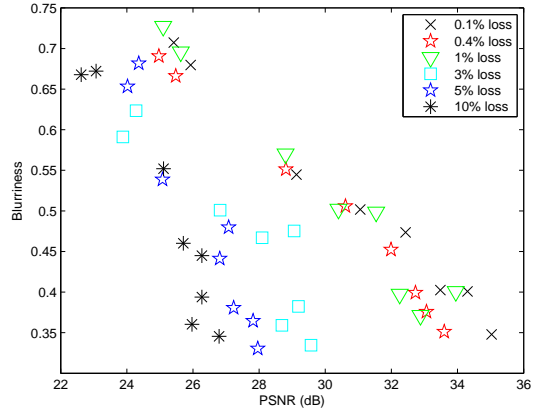
**Fig. 6.** PSNR values for the ‘paris’ video under different bit rates and packet loss rates.



**Fig. 8.** Blockiness values for the ‘paris’ video under different bit rates and PSNR values.



**Fig. 7.** Blockiness values for the ‘foreman’ video under different bit rates and PSNR values and different PLR values.



**Fig. 9.** Blurriness values for the ‘foreman’ video under different bit rates and PSNR values.

$$f_2(PLR) = -2.1 \cdot \ln(PLR) + 29.1, \quad (8)$$

where the functional forms of  $f_1$  and  $f_2$  were found by fitting the artifact metrics and PLR values to the PSNR values.

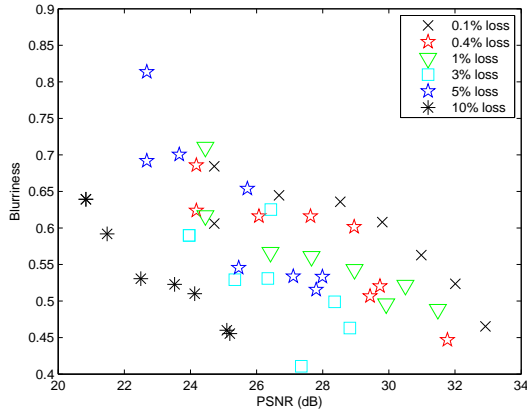
This hybrid quality metric  $Q$  takes into account the fact that, at low PLR values, the quality is well predicted by the picture metrics. But, as the PLR increases, the results for the no-reference artifact metrics start degrading because the packet losses introduce “new content” to the video in a highly nonlinear way. Because of that, we introduce the weighting factor  $\beta = PLR/\alpha$ , where  $\alpha$  works as a scaling factor that expresses the value above which the PLR becomes unbearable for the streaming video. In our tests, we found  $\alpha = 11$ .

Figure 12 depicts the results of applying the hybrid quality metric  $Q$  to the videos ‘foreman’ and ‘paris’, each compressed with target bitrates of 50k, 100k, 150k, 200k, 250k, 300k, 350k, and 400k bps, and PLR values of 0.1%, 0.4%, 1%, 3%, 5%, and 10%. The combination model has  $r =$

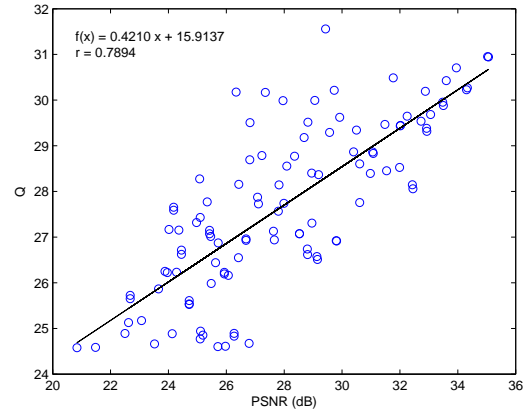
78.94%, presenting, therefore, a good performance for a no-reference quality metric.

#### 4. CONCLUSIONS

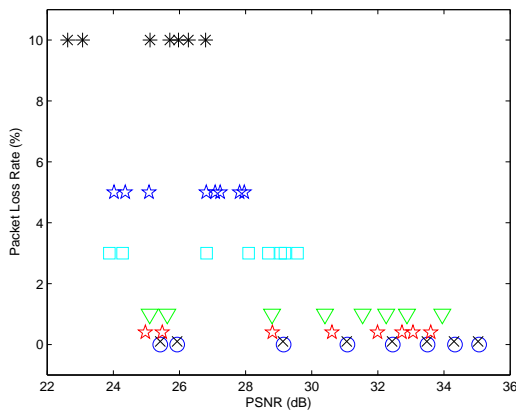
In this paper, we presented a hybrid no-reference video quality metric targeted at the transmission of videos over the Internet. The proposed metric blindly estimates the quality of videos degraded by compression and digital transmission artifacts. The metric is composed by two no-reference artifact metrics that estimate the strength of blockiness and blurriness artifacts. A combination model is used to add the packet loss rate information to the metric, eliminating the disturbance in the artifact metrics values caused by higher packet loss rates. Further studies are needed in order to better understand and characterize the interactions among the different types of artifacts and their relation to video quality.



**Fig. 10.** Blurriness values for the ‘paris’ video under different bit rates and PSNR values.



**Fig. 12.** Output values of proposed hybrid metric for videos ‘foreman’ and ‘paris’.



**Fig. 11.** Packet loss rate values for the ‘foreman’ video under different bit rates and PSNR values.

## 5. REFERENCES

- [1] ITU Recommendation BT.500-8, *Methodology for subjective assessment of the quality of television pictures*, 1998.
- [2] B. Girod, “What’s wrong with mean-squared error?,” in *Digital Images and Human Vision*, Andrew B. Watson, Ed., pp. 207–220. MIT Press, Cambridge, Massachusetts, 1993.
- [3] S. Winkler, “A perceptual distortion metric for digital color video,” in *Proc. SPIE Conference on Human Vision and Electronic Imaging*, San Jose, CA, USA, 1999, vol. 3644, pp. 175–184.
- [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] Video Quality Experts Group, “Final report from the video quality experts group in the validation of objective models of multimedia quality assessment,” Tech. Rep., <http://ftp.crc.ca/test/pub/crc/vqeg/>, 2008.
- [6] T. Vlachos, “Detection of blocking artifacts in compressed video,” *Electronics Letters*, vol. 36, no. 13, pp. 1106–1108, 2000.
- [7] 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.
- [8] G.A. Thomas, “Television motion measurement for datv and other applications,” Tech. Rep. 1987/11, BBC Res, Dept. Rep., 1987.
- [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] 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.
- [11] F. De Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, and T. Ebrahimi, “A H. 264/AVC video database for the evaluation of quality metrics,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, pp. 2430–2433.