Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Evaluation of packet loss impairment on streaming video

RUI Hua-xia, LI Chong-rong, QIU Sheng-ke
(*Network Research Center, Tsinghua University, Beijing 100084, China*)
E-mail: huaxiarui@hotmail.com; licr@cernet.edu.cn; qsk03@mails.tsinghua.edu.cn
Received Dec. 1, 2005; revision accepted Feb. 17, 2006

Abstract: Video compression technologies are essential in video streaming application because they could save a great amount of network resources. However compressed videos are also extremely sensitive to packet loss which is inevitable in today's best effort IP network. Therefore we think accurate evaluation of packet loss impairment on compressed video is very important. In this work, we develop an analytic model to describe these impairments without the reference of the original video (NR) and propose an impairment metric based on the model, which takes into account both impairment length and impairment strength. To evaluate an impaired frame or video, we design a detection and evaluation algorithm (DE algorithm) to compute the above metric value. The DE algorithm has low computational complexity and is currently being implemented in the real-time monitoring module of our HDTV over IP system. The impairment metric and DE algorithm could also be used in adaptive system or be used to compare different error concealment strategies.

Key words: Packet loss, Video streaming, MPEG-2 video, Video quality, No-Reference
doi:10.1631/jzus.2006.AS0131 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Video compression technologies have been widely used in various applications because they significantly reduce the amount of storage or bandwidth. However, they also bring the problem of perceptual quality degradation either directly or indirectly. The first category of perceptual quality degradation like blocking edge artifacts, mosaic pattern effect, blurring, etc. is directly caused by the video compression process. The second category of perceptual video quality degradation is caused by the inevitable packet loss in today's best effort IP network. Because the compression process removes both spatial and temporal redundancy of the original video, every packet is very important for the video reconstruction at the receiver side (Figs.2a and 3a show the artifacts caused by packet loss).

Accurate evaluation of the packet loss impairment on the perceptual quality of compressed video is extremely valuable because it could be used in many applications. Lu *et al*.(2002) developed an adaptive perceptual video quality control mechanism based on an application-level perceptual video quality scheme. Feamster and Balakrishnan (2002) quantified the effects of packet loss on the quality of MPEG-4 video and proved the effectiveness of their proposed adaptive system. Tao and Guerin (2004) focused on the dynamical path selection problem of streaming video from the perspective of video quality rather than the network performance. Also, some researchers used quality degradation metrics to compare error concealment strategy and system performance in the presence of errors (Chen, 1995; Zhang and Lee, 1993; Boyce and Gaglianello, 1998).

Despite their importance, methods evaluating packet loss impairment are not as many. PSNR based metrics are widely used but they do not correlate well with perceived quality measurement (Wang *et al*., 2002). Some models and evaluation methods have considered the characteristics of human visual system (HVS) in their metrics, but they often require the original video as a reference, which is not available in many real time applications. Wang *et al*.(2002) proposed a No-Reference perceptual quality assessment for JPEG compressed image, in which they used pixel value difference at each macroblock row to quantify the edges of blocking artifacts in an image. Recently, Babu *et al*.(2004) proposed an NR packet loss metric (BBPH for short) to evaluate the packet loss impairment on MPEG-2 video. The basic idea of their proposed approach is to carefully check the pixel value difference both at the horizontal boundary of each macroblock row and inside that macroblock row, point by point. If certain conditions are satisfied, that edge point is counted for the overall metric. They achieved two valuable results by their novel approach. First, their experimental results showed that metric values for unimpaired frames were very close to zero. Second, their experiment demonstrated that the metric value was proportional to the length of the artifacts. However, there are also two deficiencies in the BBPH metric. First, the BBPH metric only takes the impairment length into account, which is not appropriate because the impairment strength, which means the contrast between the impaired pixels and surrounding pixels, also affects the impairment level. Second, the evaluation process of BBPH would inevitably miscount when processing an unimpaired macroblock row just below an impaired one.

In this paper, we propose both a new NR method to evaluate the packet loss impairment on DCT based compressed video and an analytic model to explain the metric. Our method includes two essential stages, both of them described in Section 3. The first one regards the detection of impairments while the second one concerns the evaluation of those impairments. We use MPEG-2 compressed video in the discussion and experiments because MPEG-2 standard is a widely used video compression standard both in practice and in the literature.

The remainder of the paper is organized as follows. Section 2 describes the analytic model and the metric and Section 3 explains our algorithm for computing the metric. In Section 4, we present the experimental results of our algorithm. Finally, we summarize our conclusion in Section 5.

MODEL OF THE IMPAIRMENT

In our model, we consider both the impairment length and the impairment strength. There are two basic assumptions of the model. First, most unimpaired frames have smooth spatial edge, which means the differences between consecutive pixel rows do not vary too much; second, sharp edges in a frame are rarely aligned with macroblock boundaries. These two assumptions are also claimed in (Babu *et al*., 2004), and our experiments with over 100 randomly selected images also support the above assumptions.

A frame can be represented as a 2D signal. Here we denote a frame with M rows and N columns as $x(i,j)$, for $1\le i\le M$, $1\le j\le N$, $x(i,j)$ also represents the pixel value at ith row and jth column.

Let

$$\begin{gathered} cm = M/T,\ cn = N/T,\ T=16, \\ dh\_1(r,c) = \left|x(rT,c) - x(rT-1,c)\right|, \\ dh\_2(r,c) = \left|x(rT+1,c) - x(rT,c)\right|, \\ dh\_3(r,c) = \left|x(rT+2,c) - x(rT+1,c)\right|, \\ 1\le r\le cm-1,\ 1\le c\le N. \end{gathered}$$

As shown in the definition above, $dh\_1(r,c)$ is the difference inside the rth macroblock row at column c and $dh\_3(r,c)$ is the difference inside the macroblock row below at column c, while $dh\_2(r,c)$, on the other hand, is the difference at the boundary of these two macroblock rows at column c.

Then we average the above three vectors along the horizontal direction.

$$\begin{gathered} dh_1(r) = \frac{1}{N}\sum_{c=1}^{N} dh\_1(r,c), \\ dh_2(r) = \frac{1}{N}\sum_{c=1}^{N} dh\_2(r,c), \\ dh_3(r) = \frac{1}{N}\sum_{c=1}^{N} dh\_3(r,c), \end{gathered}$$

here $dh_1(r)$, $dh_2(r)$, $dh_3(r)$ denote the average value of difference at row $rT-1$, rT, $rT+1$ correspondingly. Also we use $N_1(r)$ to denote the impairment length in a macroblock row and $N_0(r)$ to denote the length of unimpaired pixels in that macroblock row. So we have

$$N_0(r) + N_1(r) = N. \tag{1}$$

To describe the strength of the difference between two consecutive rows, we use $\varepsilon_1(r)$ for the mean difference between row rT and row $rT-1$. Assuming the macroblock row starting from row rT+1 has some impairment, we use $\varepsilon_2(r)$ for the mean difference between row rT and the unimpaired part of row rT+1. For the impaired part of row rT+1, which generates a visible edge, we use δ to represent the mean difference. The model can be simply expressed in the following two equations:

$$dh_1(r) = N\varepsilon_1(r), \tag{2}$$

$$dh_2(r) = N_0\varepsilon_2(r) + N_1\delta(r). \tag{3}$$

Apparently, ε_1 and ε_2 do not generate any uncomfortable edge because they are naturally existing difference between rows. But δ generates the artifacts and hence makes the frame visually uncomfortable to watch. The hope is to derive a metric taking account of N_1 and δ. A simple metric could be:

$$E(r) = \frac{N_1(r)}{N} \cdot \frac{\delta(r) - \varepsilon_2(r)}{\varepsilon_1(r)}. \tag{4}$$

In Eq.(4), $N_1(r)/N$ measures the impairment length and $(\delta(r)-\varepsilon_2(r))/\varepsilon_1(r)$ measures the impairment strength.

Eq.(4) implicates both impairment length and impairment strength. But unfortunately it cannot be readily computed. On the other hand, based on Eqs.(2) and (3), we could easily get:

$$\begin{aligned}\hat{E}(r) &= \frac{dh_2(r) - dh_1(r)}{dh_1(r)} \\ &= \frac{N_0(r)\varepsilon_2(r) + N_1(r)\delta(r) - N\varepsilon_1(r)}{N\varepsilon_1(r)}.\end{aligned} \tag{5}$$

According to Eq.(1), the above equation can be expressed as:

$$\hat{E}(r) = E(r) + \Delta, \tag{6}$$

$$\Delta = \frac{\varepsilon_2(r) - \varepsilon_1(r)}{\varepsilon_1(r)} = \frac{\varepsilon_2(r)}{\varepsilon_1(r)} - 1.$$

The first assumption stated at the beginning of this section guarantees Δ to be very small. Thus we could compute $\hat{E}(r)$ to estimate $E(r)$ and give a reasonable evaluation of the impairments.

DETECTION AND EVALUATION ALGORITHM

With the metric deduced in Section 2, we further design a simple detection and evaluation algorithm (DE algorithm for short) to evaluate the impairment of a damaged frame. The DE algorithm runs in two stages, as the name suggests. The detection stage aims at accurate detection of impairment while the evaluation stage tries to efficiently calculate the $\hat{E}(r)$ values for each macroblock row and average them to give an assessment of the whole frame. The purpose of separating the detection process and evaluation process is to make sure that metric for those macroblock rows with no artifacts are uniformly zero and only impaired macroblock rows would generate non-zero metric values.

Detection

Because of the variation of scenes, simply calculating the difference of pixel values is not enough for accurate detection because the differences vary with the changing scenes. We solve this problem by comparing $dh_2(r)$ with both $dh_1(r)$ and $dh_3(r)$. It is also important to avoid including real edge. We solve this problem by verifying whether there are sharp edges both at the upper boundary and downside boundary because natural edges rarely occur at the boundary of a macroblock row, which is stated in the second assumption in Section 2. In fact, the possibility that an unimpaired frame has sharp edges exactly at both the upper boundary and the downside boundary of a macroblock row is extremely low. This fact guarantees the accuracy of the detection process.

To verify the existence of a sharp edge, we check whether the following condition is satisfied.

$$\begin{aligned} dh_2(r) &> normal \times \max(dh_1(r), dh_3(r)), \\ dh_2(r+1) &> normal \times \max(dh_1(r+1), dh_3(r+1)), \end{aligned} \tag{7}$$

here the parameter *normal* is chosen to be 1.5 so that normal fluctuation of the mean difference between rows is not treated as sharp edge.

Also, in order to avoid the situation when $dh_1(r)$, $dh_2(r)$ and $dh_3(r)$ are all extremely small so that Eq.(7) is accidentally met, which could happen in some uniform areas, we set a threshold for $dh_2(r)$ to eliminate that possibility. It should be noted that this threshold is not strict and therefore could apply to various frames. For our algorithm, we use 10,

$$dh_2(r) > noise, \quad noise = 6. \tag{8}$$

If Eqs.(7) and (8) are both satisfied, we assert the existence of packet loss impairment at the (r+1)th macroblock row. Currently, the parameters (*normal*, *noise*) are selected empirically to retrieve best detection results.

Evaluation

The algorithm computing the metric for a frame is as follows:

(1) Initialize variables like *cm*, *cn*, *normal*, *noise*.

(2) Compute $dh\_1(r,c)$, $dh\_2(r,c)$, $dh\_3(r,c)$.

(3) Compute average vectors $dh_1(r)$, $dh_2(r)$ and $dh_3(r)$.

(4) Start the loop from the second macroblock row. If Eq.(7) or Eq.(8) is not true, $\hat{E}(r)$=0, go to next macroblock row, otherwise, compute $\hat{E}(r)$ according to Eq.(5).

(5) Compute the overall metric of the frame $\hat{E} = \frac{1}{cm-2}\sum_{r=2}^{cm-1}\hat{E}(r)$.

The first and last macroblock rows should be evaluated additionally, because Eq.(7) could only be satisfied partially if there exist any impairments.

To calculate the impairment metric of a video, we evaluate each frame in the video separately and average them to retrieve the mean value for the video.

EXPERIMENT AND RESULTS

In this section we describe the experiment setup and results of DE algorithm.

Experiment setup

The experimental test bed is composed of three components as indicated in Fig.1.

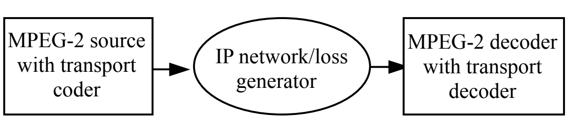


Fig.1 Experimental test bed

The first component retrieves MPEG-2 coded file and sends it to the receiver using RTP/UDP protocols. The MPEG-2 coded file used in our experiment consists of two types: the elementary streams provided by Tektronix and transport stream recoded by ourselves using JVC HDV camera. The first type was coded with bitrate=11.4 Mbps, frame rate=25 fps and frame size=704×576. The second type was coded with bitrate=20 Mbps, frame rate=30 fps and frame size=1280×720. The second component could either be a real IP network or a loss generator between the sender and receiver. In our experiment, we use CERNET (China Education and Research Network) for real IP network experiment and a software developed in our lab simulating various loss patterns for simulation. The third component decodes the received stream and stores extracted frames in the disk. We use a free MPEG-2 video stream decoder, libmpeg2, to decode video streams.

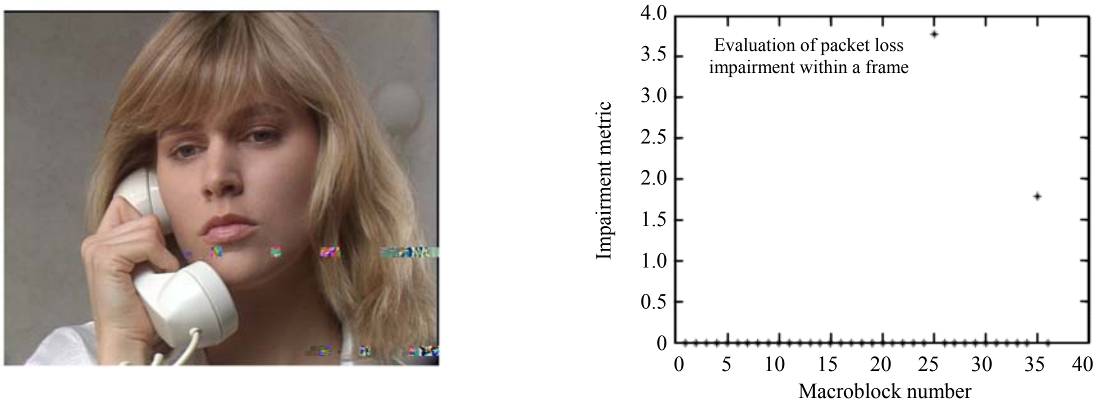
Experiment results

We sent “susi” from 166.111.203.90 to 210.25.128.189 at 7 o’clock and 10 o’clock separately. Fig.2a shows one frame extracted from 7 o’clock received video and Fig.3a shows one frame extracted from 10 o’clock received video. The impairment metric value for Fig.2a is shown in Fig.2b and the impairment metric value for Fig.3a is shown in Fig.3b. Each dot in Figs.2b and 3b corresponds to the metric value for a macroblock row. Only non-zero values indicate the existence of impairments.

In Fig.2a, there are impairments in two macroblock rows in the lower part of the frame. Correspondingly, we see two non-zero metric values in the Fig.2b. The longer impairment has a larger value, which is as expected. In Fig.3a, there are impairments in many macroblock rows. Apparently, both impairment length and impairment strength affect our subjective assessment. It should be noted that the macroblock row corresponding to the largest value is not the one with the longest impairment.

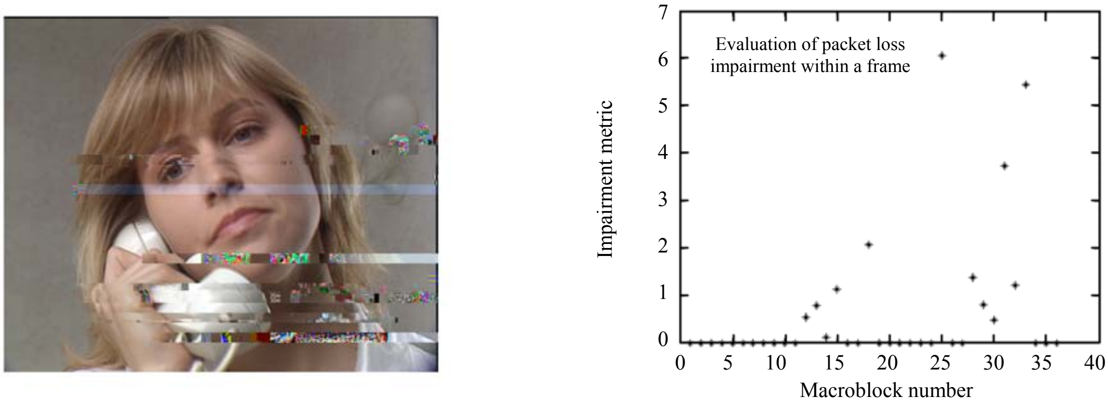
Fig.4 shows the metric values for a sequence of frames. Each dot corresponds to the impairment metric value for that frame. The sequence is extracted from an MPEG-2 HDV TS which suffered packet loss ratio of 0.1%. Here, we infer packet loss ratio to be the ratio of the number of lost packets to that of total packet sent, and the packet length used in our experiment is fixed. Because the GOP structure for this video is IBBPBB, we could see that most of the packet losses occur in the intra-coded frame and the loss effect will last for 6 frames which is the GOP size.

Fig.5 is a simulation which shows the impairment metric values of videos received from the same MPEG-2 HDV source file, but undergoing different packet loss ratio, ranging from 1% to 20%. Each dot



(a) (b)

Fig.2 (a) Impaired frame from Susi received at 7 o'clock; (b) $\hat{E}$ of 7 o'clock Susi frame



(a) (b)

Fig.3 (a) Impaired frame from Susi received at 10 o'clock; (b) $\hat{E}$ of 10 o'clock Susi frame

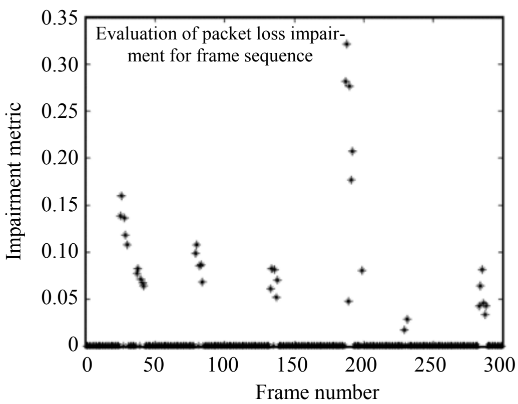


Fig.4 Impairment metric values for a frame sequence

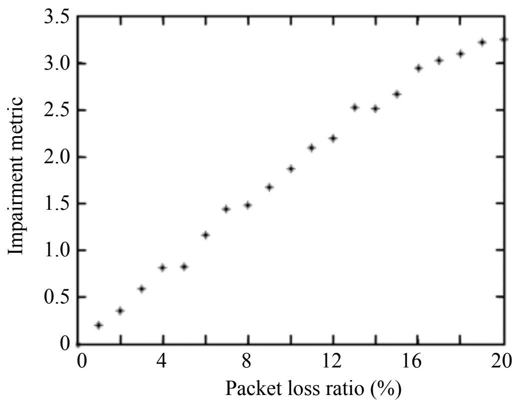


Fig.5 Impairment metric values versus PLR

in the graph represents the impairment value for a specific video. The graph suggests that impairment metric increases as packet loss ratio increases, as we expected.

CONCLUSION

In this work, we developed an analytical model to study the packet loss impairment on MPEG-2 video. Based on the model, we further proposed a No-Reference metric to evaluate the impairment and an efficient algorithm, namely the detection and evaluation algorithm (DE algorithm) to compute the metric value for a frame or a video. The algorithm gives zero value for unimpaired videos. But for an impaired video, the algorithm would give a non-zero value which takes into account both the impairment length and impairment strength. Basically, the experiment results are consistent with our experiences, but further subjective experiments are required to maximize the correlation between our impairment metric and results from subjective tests.

We are currently trying to use the proposed metric and DE algorithm for real time quality moni toring of streaming video, which is useful in adaptive system or for trouble shooting. Also, the metric could be used to compare various video streaming plans, error concealment strategies, etc.

References

Babu, R.V., Bopardikar, A.S., Perkis, A., Hillestad, O.I., 2004. No-Reference Metrics for Video Streaming Applications. International Packet Video Workshop.

Boyce, J.M., Gaglianello, R.D., 1998. Packet Loss Effects on MPEG Video Sent over the Public Internet. ACM International Multimedia Conference, p.181-190.

Chen, C.T., 1995. Error detection and concealment with an unsupervised MPEG2 video decoder. *Journal of Visual Communication and Image Representation*, **6**(3):265-279. [doi:10.1006/jvci.1995.1023]

Feamster, N., Balakrishnan, H., 2002. Packet Loss Recovery for Streaming Video. International Packet Video Workshop.

Lu, X.X., Morando, R.O., Zarki, M.E., 2002. Understanding Video Quality and Its Use in Feedback Control. International Packet Video Workshop.

Tao, S., Guerin, R., 2004. Application-Specific Path Switching: A Case Study for Streaming Video. Proceedings of ACM Multimedia.

Wang, Z., Sheikh, H.R., Bovik, A.C., 2002. No-Reference Perceptual Quality Assessment of JPEG Compressed Images. Proc. ICIP'02, **1**:477-480.

Zhang, Y.Q., Lee, X.B., 1993. Performance of MPEG Codecs in the Presence of Errors. Proc. SPIE, **2094**:1739-1749.