Video Quality Estimation for Packet Loss based on No-Reference Method

Junghyun Han, Yo-han Kim, Hyuntai Kim, and Jitae Shin

School of Information and Communication Engineering
Sungkyunkwan University, Suwon, 440-746, Korea
{kazikai, dos95, hhtsog, jtshin}@skku.edu

Abstract—The objective video quality assessment metrics can be divided into Full Reference (FR), Reduced Reference (RR), and No Reference (NR). Among them, NR quality assessment method does not require the original video sequence so it is suitable for the real-time video streaming service. This paper presents estimation method for video quality degradation due to various pattern packet losses in packet video network environment. Our proposed algorithm predicts initial error due to packet loss in each frame by using motion information of received data. Proposed method shows high Pearson correlation (more than 95%) to match actual mean square error.

Keywords—Packet Loss, Video Quality Estimation, No-Reference, Motion Vector, Mean Square Error (MSE)

I. INTRODUCTION

Recently, the development of Internet and compression codec based on real-time video streaming service’s proportion has increased rapidly. Channel error can cause reconstructed frame at the decoder to be incorrect, which leads to significant error propagation to subsequent frames. It causes video quality degradation. For video quality assessment method, video quality metrics have been developed and standardized by International Telecommunication Union (ITU) and Video Quality Experts Group (VQEG) [1] [2].

The evaluation methods of video quality are as follows. Full Reference (FR) metrics perform the distortion measure having full access to the original video sequence. Reduced Reference (RR) metrics work with some information about the original video sequence to perform quality measurements. And finally, No Reference (NR) metrics need no information of original video sequence. It requires only distorted video information [3]. The most widely used FR objective video quality metrics are Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). They are simple and quick to calculate, providing a good method to evaluate the video quality [4]. But these methods are not suitable for real-time video streaming, because they require sender-side data.

For example, MSE is defined as

$$MSE_{kth} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$ (1)

Where $y_i$ is an original video sequence and $\hat{y}_i$ is the reconstructed video sequence. In this paper, we propose an NR-based-objective quality metric that replaces the MSE metric and uses only received encoded streams to estimate reconstructed video quality.

A conventional NR method uses the NR-P (No reference-Pixel) method and NR-B (No Reference Bit-stream) methods. Primarily, blockiness analysis and blurring analysis are examples of the NR-P method [5] [6]. And the other methods are video quality assessment of bit-stream by using quantization parameter, error propagation model and error concealment method [7] [8]. But these metrics have complex formula and lower Pearson correlation [9]. So it is not suitable for real-time streaming service.

Especially this paper examines the NR method considering various packet loss patterns to estimate the MSE distortion. In order to estimate video quality, we consider H.264/AVC video streaming system over packet video network. And we assume that each frame is coded into a single packet, and the loss of a packet corresponds to the loss of an entire frame.

The organization of the paper is as follows: In the next section we describe the proposed algorithm for packet loss pattern. In section 3, we show the experimental result. Finally Section 4 presents conclusion.

II. PROPOSED ALGORITHM

In this paper, we offer objective NR video quality for video quality degradation based packet loss. Most video compression algorithms, including MPEG-1/2/4 and H.263/H.264 are based on motion compensated prediction. In video encoder case, inter-frame prediction method is used.

If packet loss occurs, then error propagation is shown in Figure 1.
If a single loss is introduced at $k$-th frame, the total MSE is given by

$$MSE_{kth} = \sum_{i=0}^{El} \left( 1 - \frac{i}{n} \right) \times MSE_{initial}.$$  \hspace{1cm} (2)

The $MSE_{kth}$ is the initial MSE of $k$-th frame’s loss. The notation $El$ is the number of frame between current loss location and next intra frame and $n$ is number of frames in a group of picture (GOP). But, if the channel environment is bad, the multiple packet loss can occur in one GOP. Therefore, we should also consider multiple packet loss case. So Eq.(2) can be changed into Eq.(3)

$$MSE_{kth} = \sum_{j=0}^{Nl} \sum_{i=0}^{El} \left( 1 - \frac{i}{n} \right) \times MSE_{initial}.$$  \hspace{1cm} (3)

The $Nl$ is the number of loss in a GOP. Eq. (3) may be used only in one GOP. So, we need an extended equation to consider total video sequence. The notation $Ng$ is the number of GOP in a video sequence. We can gather an Eq.(4) for the initial MSE of $k$-th frame’s loss

$$MSE_{kth} = \sum_{k=0}^{Ng} \sum_{j=0}^{Nl} \sum_{i=0}^{El} \left( 1 - \frac{i}{n} \right) \times MSE_{initial}.$$  \hspace{1cm} (4)

Finally, the initial MSE of $k$-th frame’s loss is obtained as Eq.(4). But, $MSE_{initial}$ value can be only estimated when the process are able to use the original video sequence. It is not NR method because, it requires initial error. So Eq.(4) is not available at the receiver.

Therefore, we need another estimation of initial MSE by using only received video sequence.

We focus on the motion information metrics. In case of video sequence, video sequence has motion vector for motion activity information in each block’s coordinates. If this motion vector value is bigger, it increases the initial value of MSE due to packet loss. In other words, fast motion video is more vulnerable to packet loss than slow motion video. The $MA$ is a key parameter to measure the motion vector. This notation is calculated as follows

$$MA = \frac{\sum_{j=0}^{m} \sum_{i=0}^{n} (mv_i^2 + mv_j^2)}{W \times H}. $$ \hspace{1cm} (5)

The $i, j$ are the coordinates of block in a frame. And the notation $W$ is the width of the video sequence and $H$ is the height of the video sequence.

We use the three test sequences of Figure 2. Sequence (a) named Mobile has small motion, (b) Foreman has middle motion, and (c) Soccer has large motion. These three sequences represent various types of video. So, those are suitable for induction between $MA$ and $MSE_{initial}$.

We performed several tests to find out the relationship between $MA$ and $MSE_{initial}$. The sequence (QCIF, 176×144) (a), (b), and (c) are encoded by using H.264 encoder(JM 12.4 [10]) with 14 types of error as pattern presented in Table 1.

<table>
<thead>
<tr>
<th>Error Pattern Setting of Figure 2’s test sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern Number</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Location of Error(P-frame)</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>

Figure 3. Relation between MA and Initial MSE

<table>
<thead>
<tr>
<th>MA and Initial MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Motion Activity</td>
</tr>
<tr>
<td>$y=963.02x-131.43$</td>
</tr>
</tbody>
</table>

Figure 3. Relation between MA and Initial MSE
Figure 3 shows that $MSE_{initial}$ and $MA$ are related in linear relationship. We can estimate the initial error from motion vector of video sequence. The estimated $MSE_{initial}$ is as follows

$$MSE_{initial} = \alpha \times MA + \beta.$$  (6)

To find the linear parameter $\alpha$ and $\beta$, we conduct recursive analysis. So, the linear parameter value is obtained as follow Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>963.02</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-131.43</td>
</tr>
</tbody>
</table>

In addition, Eq.(6) divided by the new constant to match the value of the estimated $MSE$ and actual $MSE$. The new constant is $\gamma$. The value of $\gamma$ is 3.98.

Therefore, the initial value of $MSE$ is the following approximation using Table 2 and parameter $\gamma$.

$$MSE_{initial} \approx \frac{1}{3.98} (963.02 \times MA - 131.43)$$  (7)

Finally, we can obtain our estimation $MSE$ by using Eq.(4) and Eq.(7). The estimated $MSE$ represent as only motion vector, GOP size, and packet loss pattern information of received video sequence.

$$MSE_{est} = \frac{\sum_{k=0}^{N_t} \sum_{j=0}^{N_I} \sum_{i=0}^{E_l} (1 - \frac{i}{n}) \times (241.96 \times MA - 33.01)}{15}$$  (8)

### III. SIMULATION RESULTS

In this paper, to evaluate performance of proposed algorithm, we compare the correlation between the actual $MSE$ and estimated $MSE$ value.

<table>
<thead>
<tr>
<th>GOP Size in Sequence</th>
<th>Intra frame</th>
<th>Inter frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of GOP</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Number of Frame</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>Error Pattern</td>
<td>Pattern Number</td>
<td>Location of Error</td>
</tr>
<tr>
<td>1</td>
<td>1st P-frame</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2nd P-frame</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3rd P-frame</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4th P-frame</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5th P-frame</td>
<td></td>
</tr>
</tbody>
</table>

In order to test our proposed video quality method, we use the (a) “Crew (QCIF, 176x144)” and (b) “Soccer (QCIF, 176x144)” displayed in Figure 4 are encoded using H.264 encoder (JM 12.4 [10]) with encoding parameters, as presented in Table 3. The 18 types of packet loss are introduced by JVT loss simulator [11].

$$PC = \frac{\sum (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \times \sqrt{\sum (Y_i - \bar{Y})^2}}$$  (9)

Figure 5 shows the performance comparison using test sequence (a) among actual measured $MSE$, $MSE$ by using conventional method [12], and the estimated $MSE$ by using Eq.(8). And also the case using test sequence (b) is presented in Figure 6. The Pearson correlation of these value are presented in Table 4. The Pearson correlation (PC) is the evaluation metric for objective quality assessment. Pearson correlation recommended by VQEG Hybrid / Bitstream Group test plan [9]. This $PC$ is defined as eq. (9).

The Pearson correlations of these value are presented in Table 4. The Pearson correlation is 1 in the case of an increasing linear relationship, -1 in the case of a decreasing linear relationship, If the variables are independent the Pearson correlation is 0. In other words, the Pearson correlation gives information about the degree of correlation as well as the direction of the correlation. In general, Pearson correlations above 0.80 are considered pretty high.
The conventional method’s Pearson correlation is the correlation between the actual DMOS and PSNR [12].

### Table 4. Pearson Correlation

<table>
<thead>
<tr>
<th>Video Sequence</th>
<th>Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crew</td>
<td>0.954</td>
</tr>
<tr>
<td>Soccer</td>
<td>0.983</td>
</tr>
<tr>
<td>Conventional [11]</td>
<td>0.732</td>
</tr>
</tbody>
</table>

Figure 5. Proposed Estimated MSE and Actual MSE (Crew sequence)

Figure 6. Proposed Estimated MSE and Actual MSE (Soccer sequence)

### IV. CONCLUSIONS

In this paper, we propose a NR method using motion vector and contrive objective video-quality information considering packet loss. From the experimental result, we can obtain higher Pearson correlation than conventional method [12] in comparison with the actual mean square error and its estimation value. All parameters can be extracted in decoding process. And it is a very simple and practical method. Its simplicity and easiness make this method suitable for real-time video streaming. For the future work, we plan to combine the relationship video type and video quality by the packet loss. This metrics can more accurate measurement methods for video quality.

### ACKNOWLEDGMENT

This work was supported by the National Research Foundation (NRF) grant funded by the Korea government (MEST) (2009-0071289).

### REFERENCES


[9] VQEG “Hybrid / Bit stream Group test plan”

