GOP-level Transmission Distortion Modeling for Video Streaming over Mobile Networks

Yu Han, Aidong Men, Kan Chang and Ziyi Quan
School of Information and Telecommunication Engineering
Beijing University of Posts and Telecommunications
Beijing, P. R. China
hanyu616@126.com

Abstract—A major challenge in video coding and transmission over mobile networks is that the wireless channel is error-prone and the channel resources are limited. In this work, we analyze the picture distortion caused by channel errors and the distortion propagation behavior in its subsequent frames along the motion prediction path. We propose a linear fitting approach algorithm to achieve a low complexity GOP-level transmission distortion modeling. It is a predictive modeling which allows the encoder to predict the transmission distortion before the whole GOP is compressed and transmitted. The simulation results demonstrate that the proposed modeling has low computational complexity and high accuracy. It can be used in allocating the limited channel resources optimally for mobile video applications.

Keywords—transmission distortion; predictive modeling; video coding; mobile network

I. INTRODUCTION

With the bandwidth increasing in the 3G networks and the improvement of the coding efficiency, various mobile video applications have become possible. Due to the limited bandwidth of mobile channels, video signals have to be highly compressed by efficient coding algorithms, such as H.264/AVC [1]. On the other hand, mobile communication channel is error-prone where the signal strength exhibits different kinds of fading and multipath interference effects. Therefore, packet loss can frequently occur due to transmission error of data while the compressed video data is highly sensitive to error. Because of the limited bandwidth and small picture size, the packet loss usually corresponds to whole-frame loss. More important, the transmission errors introduced in one frame will propagate to its subsequent frames along the inter prediction path and significantly degrade video presentation quality. This kind of video distortion is called transmission distortion. In order to combat the effect of packet loss, FEC-based unequal error protection is one of the efficient error control schemes used for the transmission of compressed video streaming over packet-loss networks [2], but it increases bandwidth requirement. In order to optimally allocate the limited channel resources used to perform unequal error protection, an algorithm for evaluating the importance of different frames in a group of pictures (GOP) is necessary. The authors in [3] proposed a piecewise linear-fitting approach to achieve low-complexity transmission distortion modeling with one GOP time-delay. In [4], the authors proposed a control system approach to transmission distortion modeling which allowed the encoder to predict the transmission distortion, so it could be applied to real-time wireless video communication. The authors in [5] proposed a length of error propagation (LEP) method. In [6], a model-based distortion estimation algorithm was developed.

The rest of this paper is organized as follows. We provide a formal definition of transmission distortion in the next section. In Section III, we propose the predictive modeling for GOP-level transmission distortion estimation. Simulation results and analysis are presented in Section IV. We conclude the paper in Section V.

II. TRANSMISSION DISTORTION

A typical GOP in the video sequences is composed of I-frame, P-frame and B-frame. Due to the inter prediction structure, reference pictures loss could lead error propagate along motion prediction path. Since B-frames are not used as reference pictures, we consider the frame sequences in GOP without B-frames. In this structure, different frame loss of a GOP often results in different distortion propagation behavior (as shown in Fig. 1).

Figure 1. Illustration of error propagation due to the loss of the fourth frame (up) and the eighth frame (down), respectively (the frame index in the two groups above is: 1, 15, 30)
In this work, when a frame lost in the channel, we perform frame copy (repeating the previous frame) as the error concealment method in the decoder side. In the unequal error prediction scheme, accurate evaluation of the transmission distortion in the GOP-level is the key element in assigning the channel resource optimally.

In this section, we provide a formal definition of transmission distortion [3], [4]. Let \( \hat{F}(n, i) \) and \( \tilde{F}(n, i) \) be the encoder reconstruction and decoder reconstruction of the \( i \)th pixel of the \( n \)th frame in the original video sequence, respectively. Let \( D_t(n_0) \) be the instantaneous transmission distortion including by frame copy error concealment scheme in the case of frame \( n_0 \) lost. We choose the mean square error (MSE) between its encoder and decoder reconstructions to calculate \( D_t(n_0) \) (as shown in Eq. (1))

\[
D_t(n_0) = \frac{1}{N_p} \sum_{i=0}^{N_p-1} [\hat{F}(n_0, i) - \tilde{F}(n_0, i)]^2 ,
\]

where \( N_p \) is the total number of pixels in a frame. Due to the inter prediction structure, reference pictures loss could lead error propagate along motion prediction path. We define the GOP-level transmission distortion caused by frame \( n_0 \) loss to be \( D_G(n_0) \) as follows:

\[
D_G(n_0) = \sum_{n=n_0}^{N_G} D_t(n) = \frac{1}{N_p} \sum_{n=n_0}^{N_G} \sum_{i=0}^{N_p-1} [\hat{F}(n, i) - \tilde{F}(n, i)]^2 ,
\]

where \( N_G \) is the number of frames in a GOP. As shown in Fig. 4, losing different frames in the same GOP result in different GOP-level transmission distortion.

III. PREDICTIVE MODELING FOR GOP-LEVEL TRANSMISSION DISTORTION ESTIMATION

A. Distortion Propagation Behavior Caused by Whole Frame Loss

The authors in [8] demonstrated that the transmission distortion had an exponential fading behavior. The reason is that few mistakes can be smoothed out by a large number of error-free blocks using subpixel interpolation [3]. In this work, we suppose that packet loss corresponds to one whole frame loss because of the low bit rate and small picture size in mobile video applications. The high error rate invalidates the filtering efficiency of interpolation. It can be seen from Fig. 2, the transmission distortion has an approximate linear fading behavior.

![Figure 2. The propagation behavior of whole frame loss, and the linear fitting for the transmission distortion](image)

We simulate and analyze the propagation behavior of different test sequences by using H.264/AVC reference software JM 15.1 [7]. We introduce a frame loss at the ninth frame, and frame copy error concealment method is used. As shown in Fig. 2, we plot the transmission distortion as a function of frame index, and use a linear function to approximate the distortion propagation behavior. We use the proportion of the trapezium to approach the transmission distortion from frame \( n_0 \) to the end of the GOP. Let \( a, b \) and \( c \) denotes the two sides and the height of the trapezium, respectively. The estimated GOP-level distortion can be calculated by Eq. (3)

\[
\hat{D}_G(n_0) = (a + b) \frac{c}{2} , \text{ with } a = D_t(n_0) , \quad b = f(a) .
\]

As discussed above, in order to estimate GOP-level distortion, we have to calculate the value of \( b \). The next section we discusses how to predict \( b \) from \( D_t(n_0) \).

B. Proposed Predictive Modeling

The distortion propagation is included in by the inter prediction coding scheme when errors occur in reference frames. Although the motion compensation behavior is highly correlated to the video contents, the frame copy error concealment scheme makes the instantaneous transmission distortion connects with the video contents closely. Obviously, the more instantaneous transmission distortion occurs in the reference frames the more errors will propagate into the subsequences. Base on the above analysis, it is reasonable to define a relationship between \( b \) and \( D_t(n_0) \) : \( b = f(D_t(n_0)) \). As shown in Fig. 2, the transmission distortion fading behavior is approximate linear fading, so the slope of the approach linear (\( d \) as shown in Fig. 2) expresses the speed of distortion fading. It can be defined as follows:
where \( GOP_{\text{size}} \) is the size of GOP (in this work, it equal to 30). \( Slope(n_0) \) is the slope of approach linear fading in the case of frame \( n_0 \) loss. We assume that predictive modeling is accuracy, then, we get the following relationship:

\[
\hat{D}_c(n_0) = \frac{D_s(n_0) - b}{GOP_{\text{size}} - n_0},
\]

(4)

\[
b = \frac{2}{c} D_c(n_0) - D_s(n_0).
\]

(5)

From Eq. (4) and Eq. (5), we can achieve Eq. (6) as follows:

\[
Slope(n_0) = \frac{2D_s(n_0) - 2}{GOP_{\text{size}} - n_0}.
\]

(6)

We experiment on different kinds of sequences, in order to analyze the relationship between \( D_s(n_0) \) and \( Slope(n_0) \) through Eq. (6). It can be seen from Fig. 3, the envelopes of \( Slope(n_0) \) are similar to instantaneous transmission distortion \( D_s(n_0) \)'s in all of the test sequences. Although the \( Slope(n_0) \) and \( D_s(n_0) \)'s envelopes have large differences at the end of each GOP, the effect is very small since the error propagation distance is short.

\[
\text{Figure 3. Envelopes comparison between \( Slope(n_0) \) and instantaneous transmission distortion \( D_s(n_0) \)}.
\]

In order to make an optimal tradeoff between prediction accuracy and coding complexity, we use Eq. (7) to calculate the estimation of \( Slope(n_0) \). It is easy to implement with shifter. According to the simulation results in the next section, it is accuracy and robust.

\[
\hat{Slope}(n_0) = 0.025D_s(n_0).
\]

(7)

Taking Eq. (7) into Eq. (3), we achieve the predictive modeling as follows:

\[
\hat{D}_c(n_0) = (a + b) \frac{c}{2}
\]

\[
= (D_s(n_0) + (D_s(n_0) - \hat{Slope}(GOP_{\text{size}} - n_0))) \frac{c}{2}
\]

\[
= D_s(n_0)(2 - 0.025(GOP_{\text{size}} - n_0)) \frac{c}{2}.
\]

(8)

This proposed algorithm has low computational cost and low time-delay. Because of its predictive scheme, it does not need frame level distortion data accumulating operations (which requires \( \sum_{n=1}^{GOP_{\text{size}}} n = 465 \) times, whereas, the proposed algorithm only need 30 times). It is easy to be implemented with adder and shifter.

IV. SIMULATION RESULTS

We perform simulations on H.264/AVC reference software JM 15.1 [7]. The test sequences used for the experiments include Akiyo, Carphone, Coastguard, Salesman and Foreman which have a wide spectrum of scene characteristics. The GOP is composed of one I-frame (QP=28) followed by “\( GOP_{\text{size}} - 1\)” P-frames (QP=29, and one previous frame is used as reference).
As shown in Fig. 4, although the inter frame intra coding, deblending filters and non-integer motion compensation make the estimation of error propagation behavior become difficult, the proposed predictive modeling is still accuracy and robust.

Since the most important application of transmission distortion estimation is the assessment of importance for different frames in a GOP, we use the Eq. (9) as specified in [3] to measure the relative importance judgment error.

\[
E = \frac{\sum_{n=1}^{GOP_{\text{size}}} e(n) \times 100\%}{GOP_{\text{size}}}.
\]  

(9)

Here \(e(n)\) is the error flag for frame \(n\), if the estimated priority of frame \(n\) equals to its priority assigned by actual transmission distortion, \(e(n)=0\); otherwise, \(e(n)=1\).

We take experiment to compare the relative importance judgment error of different transmission distortion estimation modeling. In this experiment, the frames in a GOP are signed into two classes: the first 20% frames with the highest GOP-level transmission distortion are assigned to the premium
class, and the rest is signed as normal class. Scheme-1 is the LEP method used in [8]. Scheme-2 uses the model-based distortion estimation proposed in [9]. Scheme-3 is the fitting scheme used in [3]. More details can be found in [3]. Scheme-4 is our proposed algorithm.

<table>
<thead>
<tr>
<th>Video Sequence</th>
<th>Relative Importance Judgment Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scheme-1</td>
</tr>
<tr>
<td>Akiyo</td>
<td>29</td>
</tr>
<tr>
<td>Carphone</td>
<td>16</td>
</tr>
<tr>
<td>Foreman</td>
<td>16</td>
</tr>
<tr>
<td>Coastguard</td>
<td>33</td>
</tr>
<tr>
<td>Salesman</td>
<td>11</td>
</tr>
</tbody>
</table>

We can see from Table I that our proposed model has significant improvement comparing with the other models. Beside the accuracy, the proposed model has the lowest complexity. Although our proposed model is worse than scheme-3 in two test sequences, the proposed algorithm is predictive modeling with lower complexity which allows predicting the transmission distortion before encoding the whole GOP for real time application. So our proposed predictive modeling is accuracy and robust in mobile video applications.

V. CONCLUSION

In this paper, we have provided a formal definition of transmission distortion at first. Based on the study on the propagation behavior of whole frame loss transmission distortion, we have developed a low complexity GOP-level transmission distortion modeling. According to the simulation results, the proposed predictive modeling is accuracy and robust. It can be incorporated with channel resources allocation algorithm to improve system performance for real-time mobile video communications.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support by the National Science Foundation of China under grant number 60672087.

REFERENCES


