This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

Transmission Distortion Modeling for Wireless Video Communication

Janak U. Dani and Zhihai He
Dept. of Electrical and Computer Engineering
University of Missouri, Columbia, MO 65211
HeZhi@missouri.edu

Xiong Hongkai
Institute of Image Communication and
Information Processing, Shanghai Jiaotong
University, Shanghai 200030, P. R. China

***Abstract*— In video transmission over wireless networks, the channel is error-prone and the compressed video data is highly sensitive to errors. The transmission errors will cause decoding failure and distort the pictures. This type of picture distortion is called transmission distortion. In this work, we develop a predictive approach for transmission modeling so that the encoder is able to estimate the average transmission distortion of the decoded video data. Our simulation results demonstrate that the proposed model is robust and accurate. Using this type of predictive transmission distortion model, the encoder is able to intelligently allocate its resources, such as bits and energy, to maximize the decoded video quality at receiver side.**

I. Introduction

Video communication over wireless networks is increasingly the technology of choice for a wide range of important applications, including remote surveillance, security monitoring, smart home, environmental tracking, and battlefield intelligence. In wireless communication, the channel is time-varying and error-prone [1]. During video transmission over wireless channels, each video packet can be received correctly, received with errors, or lost by the network [2]. If the packet is received with errors, the error bits will cause decoding failure, and the video decoder has to discard the data in the packet [3]. The discarded video data and the lost video packets will cause signal distortion in the decoded video frame. More importantly, the decoding error introduced in one frame will propagate to its subsequent frames along motion prediction paths because the video data is compressed with motion prediction [3], [5], [6]. Because of error propagation, the transmission errors will accumulate over time and significantly degrade the video presentation quality, as illustrated in Fig. 1.

To model the impact of transmission errors on the video quality, an error sensitivity metric is often used for performance optimization in wireless video communication [4]. A statistical simulation approach is proposed in [6], where the decoding and error concealment behavior of the receiver is simulated at the encoder side to estimate the transmission distortion of the decoded video in a statistical average sense. This type of estimation schemes have high computational complexity and implementation cost. In addition, it does not lead to an analytic model for transmission distortion, which is however needed for performance optimization and resource allocation. In [5], the behavior of transmission errors has been extensively studied, and an empirical transmission distortion model is proposed. To estimate the model parameters, offline measurements of the transmission distortion are used. This type of offline transmission distortion modeling is often not allowed in real-time wireless video communication.

Fig. 1. The error propagation and accumulation.

In this work, our objective is to study the behavior of transmission errors in an end-to-end wireless video communication system and develop an analytic model for the transmission distortion. Using this model, the sender (video encoder) should be able to predict the average transmission distortion in the decoded video data. Based on the transmission distortion model, the encoder is able to perform optimum resource allocation to maximize the video encoding and transmission performance.

The rest of the paper is organized as follows. In Section II, we outline the proposed approach for transmission distortion modeling. In Section III, we study the fading behavior of the transmission distortion, and propose an exponential model to describe this fading behavior. In Sections IV and V, we estimate the parameters of the exponential model. Section VI presents the transmission distortion estimation scheme. The simulation results are given in Section VII. Section VIII concludes the paper.

II. Overview of the Proposed Approach

Let $F(n, i)$ be the i-th pixel of the n-th frame in the original video sequence, and $\hat{F}(n, i)$ be the corresponding reconstructed pixel at the encoder. We denote the reconstructed pixel at the receiver end as $\tilde{F}(n,i)$. By definition, the transmission distortion is

$$D_t(n) = E\left\{[\hat{F}(n,i) - \tilde{F}(n,i)]^2\right\}. \quad (1)$$

If there is no transmission errors, the decoder and encoder reconstructions should be the same, i.e.,

$$\hat{F}(n,i) = \tilde{F}(n,i), \quad (2)$$

and the transmission distortion $D_t(n)$ becomes zero. In the following, we describe our basic idea for transmission distortion modeling.

It is well known that the behavior of a discrete-time linear time-invariant (LTI) system is fully characterized by its unit impulse response $h(n)$. For an arbitrary input $x(n)$, the system output $y(n)$ is given by

$$y(n) = \sum_{k=0}^{\infty} x(n-k)h(k). \quad (3)$$

From the transmission distortion modeling perspective, we can view the wireless video communication as a "control system" with the transmission error $\mathcal{E}(n)$ as the input, and the transmission distortion $D_t(n)$ as the system output, as illustrated in Fig. 2. Here, $\mathcal{E}(n)$ represents the amount of transmission errors introduced to the video bit stream at frame n. The purpose of transmission distortion modeling is to analyze the system behavior such that we can predict $D_t(n)$ from $\mathcal{E}(n)$. To this end, as suggested by (3), we need to study the "impulse transmission distortion" behavior of the wireless video communication system. If a unit transmission error, denoted by $\delta_e(n)$, is introduced to the video bit stream at frame n_0, and the rest frames are error free, this error will create a picture distortion at frame n_0 and propagate to its subsequent frames, as illustrated in Fig. 2. We call this distortion as the impulse transmission distortion and denote it by $\mathcal{H}(n)$. We can view $\mathcal{H}(n)$ as the system response to input $\delta_e(n)$. If we assume that the wireless video system is LTI, and the transmission error at each frame is given by $\mathcal{E}(n)$, then the transmission distortion $D_t(n)$ is given by

$$D_t(n) = \sum_{k=0}^{\infty} \mathcal{E}(n-k)\mathcal{H}(k). \quad (4)$$

Obviously, to justify this control system approach and use (4) to predict the transmission distortion, we need to develop a scheme to estimate the impulse transmission distortion $\mathcal{H}(k)$. In addition, we need to demonstrate the effectiveness of the estimate scheme in (4). It should be noted that (4) only describes our basic concept and approach for transmission distortion modeling. The physical meanings of the elements in the formulation may be different from those in the control system.

III. The Fading Behavior of the Impulse Transmission Distortion

In this section, we study the fading behavior of transmission distortion and develop a scheme to estimate the impulse transmission distortion function $\mathcal{H}(n)$. The following CIF (352×288) video sequences are used during our experimental study: "Foreman", "NBA", "Football", "Coastguard", "Flowergarden", and "Stefan". These sequences have a wide range of scene characteristics. The encoding structure is an Intra frame followed by a sequence of P frames. We introduce a transmission error, for example 8% packet loss, at frame $n_0 = 2$, and the rest frames are error free. Here, the frame number 2 is arbitrarily chosen. Let

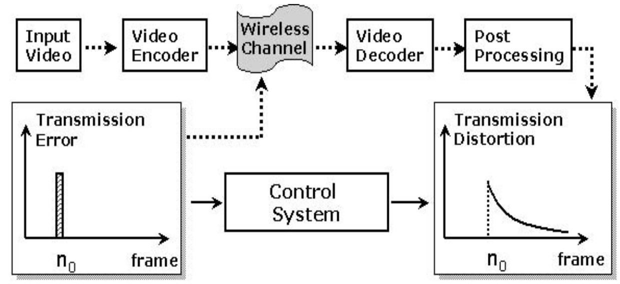

Fig. 2. A control system approach to transmission distortion modeling.

$$\bar{D}_t(n) = \frac{D_t(n)}{D_t(n_0)}, \quad (5)$$

which is the normalized transmission distortion. In Fig. 3, we plot the normalized transmission distortion $\bar{D}_t(n)$ as a function of the frame index (or time).

If we approximate this fading behavior with an exponential function, then the impulse transmission distortion can be written as

$$\mathcal{H}(n) = D_t(n_0)e^{-\alpha(n-n_0)}, \quad (6)$$

where α is called the fading factor, and $D_t(n_0)$ is the amount transmission distortion introduced at frame n_0. We call $D_t(n_0)$ as instant transmission distortion. Fig. 4 shows that this approximation is fairly close. To use the model in (6) to predict the impulse transmission distortion $\mathcal{H}(n)$, we need to estimate the fading factor α and the instant transmission distortion $D_t(n_0)$, which will be discussed in the following sections.

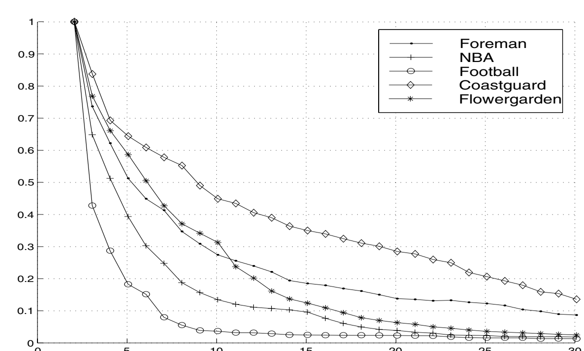

Fig. 3. The fading behavior of the impulse transmission distortion. The horizontal axis represents the frame index, and the vertical axis represents the normalized transmission distortion.

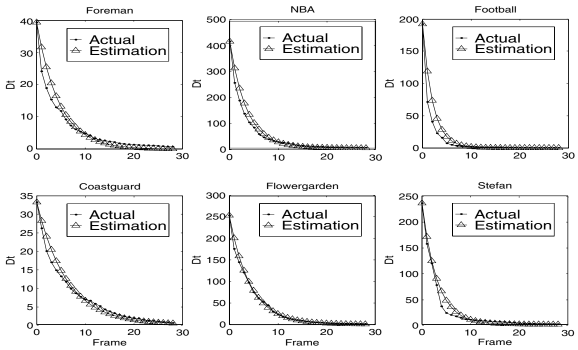

Fig. 4. Estimation of the impulse transmission distortion.

IV. Estimating the Fading Factor

In this section, we introduce an important measure, called motion reference ratio (MRR). We find that there is a very strong correlation between MRR and the fading factor. Based on this correlation, we develop a model to estimate the fading factor. The MRR is defined as the faction of pixels in frame $n-1$ which are used for motion prediction of frame n, and denoted by $\mathcal{M}(n)$. Because of motion prediction, the transmission error in frame $n-1$ will have a chance to propagate into frame n. Note that the larger $\mathcal{M}(n)$ is, more pixels in frame $n-1$ are used for motion prediction, therefore, more transmission errors in frame $n-1$ will propagate into frame n. On the other hand, the smaller $\mathcal{M}(n)$ is, less transmission errors in frame $n-1$ will propagate into frame n, which implies the transmission errors in frame $n-1$ will fade away more quickly. Therefore, we can see that there is a direct relationship between the MRR $\mathcal{M}(n)$ and the fading factor α.

In Fig. 5, we plot the fading factor and the MRR values for each test. It can be seen that $\alpha(n)$ and $\mathcal{M}(n)$ are highly correlated, which can be modeled as follows:

$$\alpha(n) = \frac{\kappa_0}{\bar{\mathcal{M}}(n)} + \kappa_1. \tag{7}$$

From our simulations, we have obtained that $\kappa_0 = 0.91$ and $\kappa_1 = -0.86$. In this work, we use (7) to predict the fading factor α of the impulse transmission distortion $\mathcal{H}(n)$.

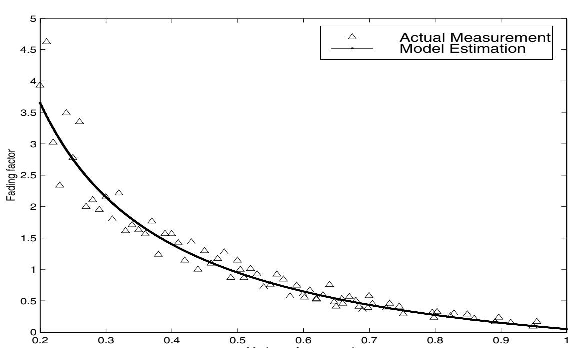

Fig. 5. Estimation of the fading factor using the motion reference ratio.

V. Estimating the Instant Transmission Distortion

In this section, we propose a scheme to estimate the instant transmission distortion $D_t(n_0)$. If there is no packet loss, $D_t(n_0) = 0$, and the reconstructed frame at the decoder is exactly the encoder reconstruction $\hat{F}(n_0, i)$. If there is a packet loss, in our error concealment scheme, the pixel in the previous decoded frame $\hat{F}(n_0-1, i)$ is used as reconstruction. Therefore, if the packet loss ratio is p, the corresponding distortion should be

$$D_t(n_0) = p \cdot RFD(n_0, n_0 - 1). \tag{8}$$

where

$$RFD(n_0, n_0-1) = E\left\{[\hat{F}(n_0, i) - \hat{F}(n_0-1, i)]^2\right\}, \tag{9}$$

representing the MSE between the reconstructed frames. Note that, at the encoder side, $RFD(n_0, n_0-1)$ can be easily obtained.

VI. Transmission Distortion Analysis

In the previous sections, we have studied the fading behavior of the transmission distortion, and developed an exponential model for the impulse transmission distortion $\mathcal{H}(n)$, as shown in (6). Once we have obtained the impulse transmission distortion, based on the linear system approach as described in (4), we can estimate the transmission distortion $D_t(n)$ for the generic case where the transmission errors are introduced in all frames. The amount of transmission errors in frame n is $\mathcal{E}(n)$. Here, the transmission error is in form of packet loss. Suppose $\mathcal{E}(n) = p(n)$, where $p(n)$ is the packet loss ratio in frame n. For a time-varying channel, $p(n)$ changes over the time index n. According to (8), if a transmission error at frame n with a packet loss ratio of $p(n)$ will cause distortion $p(n) \cdot RFD(n, n-1)$. This transmission distortion will propagate to its subsequent frames. Let $D_t(n \to n+k)$ be the amount of distortion which is introduced at frame n and propagated to frame $n+k$. According to (6), we have

$$D_t(n \to n+1) = p(n) \cdot RFD(n, n-1)e^{-\alpha(n)}, \tag{10}$$

and in general

$$D_t(n \to n+k) = p(n) \cdot RFD(n, n-1)e^{-\sum_{i=0}^{k-1} \alpha(n+i)}. \tag{11}$$

Therefore, the transmission distortion at frame n should be the summation of all transmission distortion that is introduced in its previous frames and propagated to frame n. In other words,

$$D_t(n) = \sum_{k=0}^{n-1} D_t(k \to n) \tag{12}$$

$$= \sum_{k=0}^{n-1} p(k) \cdot RFD(k, k-1)e^{-\sum_{i=k}^{n-1} \alpha(i)} \tag{13}$$

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

VII. Experimental Results

In this section, we present experimental results to demonstrate the efficiency of the proposed transmission distortion model. Figure 4 shows the packet loss ratio statistics of the time-varying channel which obtained from a multi-hop wireless ad hoc network simulation. Figure 5 shows the MRR value of each frame of the test videos. Each simulation is run for 25 times to obtain the average transmission distortion. The actual transmission distortion and the estimated one with the proposed transmission distortion model are shown in Figure 6. It can be seen that the transmission distortion model is quite accurate and robust, given the time-varying behavior of the wireless channel.

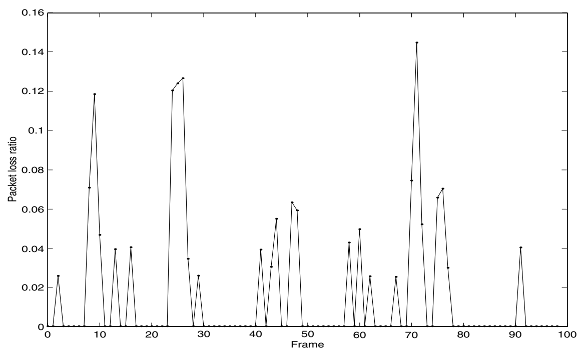

Fig. 6. The packet loss ratio at each frame for a time-varying channel.

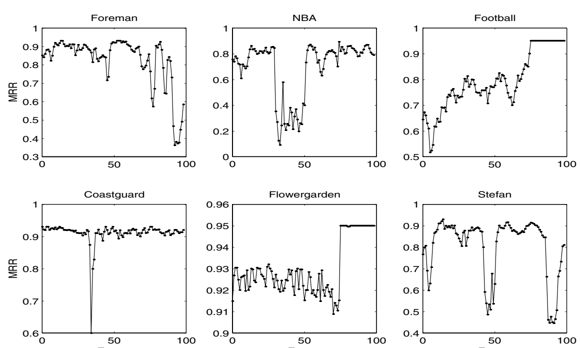

Fig. 7. The motion reference ratio of each frame.

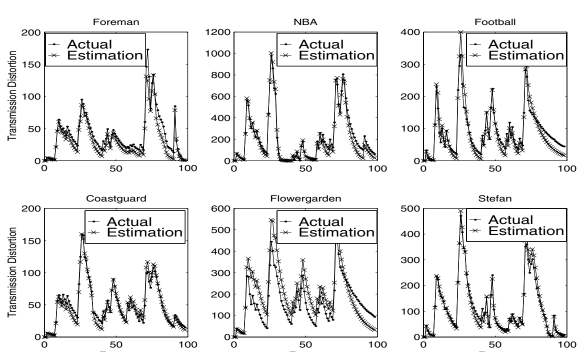

Fig. 8. Estimation of the transmission distortion for the time-varying channel.

VIII. Conclusion

In this work, we have analyzed the fading behavior of the impulse transmission errors, and developed an exponential model to predict this fading behavior. We have found that the fading factor is highly correlated with the motion reference ratio of the video sequence, and develop a model to predict the fading factor. Based on the exponential model, we develop a predictive model for the transmission distortion, with which the encoder is able to predict the transmission distortion at the decoder side. Clearly, this type of model has important applications in resource allocation and performance optimization in wireless video communication.

References

[1] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, New Jersey, 1996.

[2] D. Wu, T. Hou, Y.-Q. Zhang, "Transporting Real-time Video over the Internet: Challenges and Approaches," *Proceedings of the IEEE*, vol. 88, no. 12, pp. 1855 -1875, Dec. 2000.

[3] Zhihai He, Jianfei Cai, and Chang Wen Chen, "Joint Source Channel Rate-Distortion Analysis for Adaptive Mode Selection and Rate Control in Wireless Video Coding," *IEEE Transactions on Circuits and System on Video Technology, special issue on wireless video*, vol. 12, pp. 511 - 523, June 2002.

[4] J. Y. Liao and J. Villasenor, "Adaptive intra block update for robust transmission of H.263," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, pp. 30 – 35, February 2000.

[5] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1012 – 1032, June 2000.

[6] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 966 – 976, June 2000.

