RECURSIVE DISTORTION ESTIMATION FOR HYBRID DIGITAL-ANALOG VIDEO TRANSMISSION

Jing Zhang[†], *Anhong Wang*[†⋆], *Jie Liang*[††], *Senior Member, IEEE*

[†]Institute of Digital Multimedia and Communication,
Taiyuan University of Science and Technology, Taiyuan, 030024, China.
[††] Engineering Science, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada.
Email: jingz_tyust@163.com, ⋆ corresponding author: wah_ty@163.com, jiel@sfu.ca.

ABSTRACT

Recently, hybrid digital-analog (HDA) video transmission scheme has shown advantages in avoiding the *cliff effect*. However, most HDA schemes assume perfect transmission of digital signals, which is hardly the case in practice. This paper proposes a scheme named RDE-HDA that recursively estimates the decoder-side distortion from the encoder-side for HDA system, and both transmission error and superposition process of digital and analog parts are taken into consideration. Therefore, our method does not require the digital part of the HDA output to be transmitted losslessly, making the HDA more practical. We derive the closed-form expression of the recursive distortion estimation for HDA system. The accuracy of our method is verified by simulation results.

Index Terms— Distortion estimation, hybrid digital-analog, superposed modulation, video coding, video transmission.

1. INTRODUCTION

Wireless video transmission has been increasingly popular in recent years. Conventional digital communication systems suffer from the quality *saturation effect* and *cliff effect* [1]. In [2],[3], an analog-like transmission method called Softcast is proposed to overcome the disadvantage of conventional digital schemes. Softcast directly transmits the linear transform of the video signal by amplitude modulation without using quantization and entropy coding. Therefore the video quality is linearly related to the channel signal-to-noise ratio (CSNR). However, the compression efficiency of Softcast is quite low, which leads to lower video quality than digital schemes at the same channel bandwidth.

Recently, HDA transmission scheme is proposed to integrate the advantages of both digital and analog methods [4],[5]. The basic idea of HDA transmission is to transmit the quantization errors of digital signals in an analog mode by the aforementioned Softcast, while the digital part is dealt with by the conventional digital source coding. Then, the superposition of the digital part and analog part is transmitted over the noisy channel. Note that, the digital part is assumed to be decoded losslessly under a required CSNR.

It should be noted that the assumption of perfect transmission for the digital part poses strong constraints for HDA. For instance, the bit-error-rate (BER) of digital transmission is required to be lower than a threshold such as 10^{-8} in [5]. To meet this requirement on BER, HDA usually adopts high-rate channel coding, strong modulation, retransmission, or large power for the digital part, but these approaches reduce the efficiency of the system.

In many joint source-channel coding schemes, the goal is to reduce the end-to-end distortion (EED) [6]. In [7-9], a recursive optimal per-pixel estimation (ROPE) method is proposed to estimate the decoder-side distortion from the encoder side. Many methods have been developed to generalize the ROPE algorithm. For example, in [11], some approximations are proposed to reduce the complexity of ROPE when subpixel motion estimation is used. A fast recursive algorithm to estimate the decoder-side distortion of each frame with packet loss is also proposed in [6]. In [10], the ROPE method is generalized to estimate the decoder-side distortion of the synthesized virtual views in multiview video transmission.

In this paper, we generalize the ROPE algorithm to HDA, and propose a recursive distortion estimation method for HDA-based video transmission, abbreviated as RDE-HDA. First, we consider the distortion due to the superposition adopted in HDA system; second, we derive the closed-form expression of the video distortion by considering both transmission error and superposition error. To our best knowledge, this is the first paper on the recursive end-to-end distortion estimation of the HDA scheme. Experimental results show that the distortion estimation of our scheme is very accurate.

2. RELATE WORK

2.1. HDA system review

Fig. 1 shows the framework of HDA, consisting of the digital and analog parts in both the encoder and decoder. The

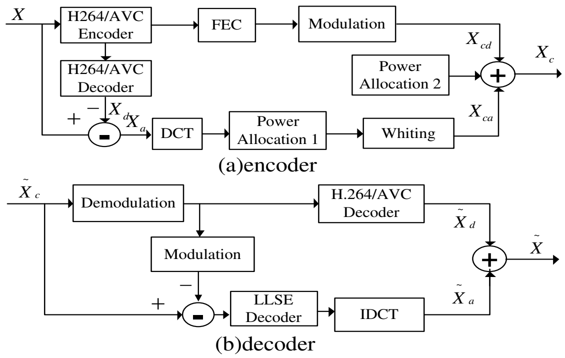


Fig. 1. The framework of HDA system.

video is first processed by H.264/AVC video encoding, forward error correction (FEC), modulation (such as BPSK), and obtain the modulated digital signal $\mathbf{X}_{\boldsymbol{cd}}$. Next, the residual $\mathbf{X}_{\boldsymbol{a}}$ between the original signal $\mathbf{X}$ and the error-free reconstructed digital signal $\mathbf{X}_{\boldsymbol{d}}$ is obtained. It is then processed following the idea of Softcast. Softcast operates on small blocks, and each frame is divided into blocks of 8×8 pixels in this paper. First, the DCT transform is performed to remove the spatial redundancy of video frame and obtain the DCT coefficients for blocks. To simplify the notation, the DCT output of *i-th* block in frame *n* is represented by a 64×1 vector $\mathbf{X}_{\boldsymbol{a}}(i, n)$. Then, a power allocation algorithm is employed to scale $\mathbf{X}_{\boldsymbol{a}}(i, n)$. These scaled vectors are assigned to packets and multiplied by a Hadamard transform matrix $\boldsymbol{H}$ during the process of whitening to get equal power packets. Thus, the transmitted analog signal is

$$\mathbf{X}_{\boldsymbol{ca}}(i, n) = \boldsymbol{HG}\mathbf{X}_{\boldsymbol{a}}(i, n) \tag{1}$$

where $\boldsymbol{G}$ is a diagonal matrix and the *k-th* diagonal entry g_k is obtained by minimizing the reconstruction distortion [2]:

$$g_k = \sqrt{\frac{P_a}{\sqrt{\lambda_k} \sum_j \sqrt{\lambda_j}}} \tag{2}$$

where λ_j and P_a are the *j-th* diagonal element of the DCT coefficient covariance matrix and the total power of each block.

The transmitted signal $\mathbf{X}_{\boldsymbol{c}}(i, n)$ is the sum of analog signal $\mathbf{X}_{\boldsymbol{ca}}(i, n)$ and the modulated digital signal $\mathbf{X}_{\boldsymbol{cd}}(i, n)$, i.e.,

$$\mathbf{X}_{\boldsymbol{c}}(i, n) = \mathbf{X}_{\boldsymbol{ca}}(i, n) + \mathbf{X}_{\boldsymbol{cd}}(i, n) \tag{3}$$

After that, the superposition signals are transmitting into wireless channel. We assume OFDM system is used. Therefore the channel can be converted into some sub-channels, and the noise in each sub-channel can be considered as additive white Gaussian noise (AWGN), which follows normal distribution, i.e., $N \sim N(0, \sigma^2)$. The received signal is

$$\widetilde{\mathbf{X}}_{\boldsymbol{c}}(i, n) = \mathbf{X}_{\boldsymbol{c}}(i, n) + \mathbf{N}(i, n) \tag{4}$$

At the decoder, the received superposed signals $\widetilde{\mathbf{X}}_{\boldsymbol{c}}$ is firstly demodulated to obtain the digital signal by treating the analog signal as noise. Then, the received digital signal is deducted from the superposed signals to obtain the analog symbols. For the analog signal, the decoder adopts a linear least square estimator (LLSE) to remove the noise, before applying inverse DCT to reconstruct the analog signal. Finally, the source sequence is reconstructed via the superposition of the decoded digital signal and analog signal, as shown in (5).

$$\widetilde{\mathbf{X}}(i, n) = \widetilde{\mathbf{X}}_{\boldsymbol{a}}(i, n) + \widetilde{\mathbf{X}}_{\boldsymbol{d}}(i, n) \tag{5}$$

2.2. Overview of the ROPE scheme

In single-view video transmission, the ROPE method aims to estimate the decoder-side distortion due to transmission errors from the encoder-side. Let $X(m, i, n)$ denotes the known *m-th* pixel in the *i-th* block of the *n-th* uncompressed frame, $\widehat{X}(m, i, n)$ and $\widetilde{X}(m, i, n)$ represent the pixel value reconstructed by the encoder and decoder, respectively. Using the mean squared error (MSE) as distortion metric, the expected decoder-side distortion

$$\begin{aligned} ED(m, i, n) =& E\{(X(m, i, n) - \tilde{X}(m, i, n))^2\} \\ =& X(m, i, n)^2 - 2X(m, i, n)E\{\tilde{X}(m, i, n)\} \\ &+ E\{\tilde{X}(m, i, n)^2\} \end{aligned} \tag{6}$$

This shows that the expected distortion is characterized by the first and second moments of $\widetilde{X}(m, i, n)$, which can be recursively obtained, depending on whether the pixel is intra-coded or inter-coded.

First, we consider intra-coded pixels. If a pixel is received correctly, the decoder reconstruction is the same as the encoder reconstruction, i.e., $\tilde{X}(m, i, n) = \widehat{X}(m, i, n)$. Otherwise, the ROPE method uses the co-located pixel from the previously reconstructed frame as the reconstruction of the missing pixel, i.e., $\tilde{X}(m, i, n) = \tilde{X}(m, i, n-1)$. Thus, the two moments ($k = 1, 2$) can be written as

$$\begin{aligned} E\{\tilde{X}(m, i, n)^k\} =& (1-p)\hat{X}(m, i, n)^k \\ &+ pE\{\tilde{X}(m, i, n-1)^k\} \end{aligned} \tag{7}$$

where *p* is the packet loss rate of the frame.

For a pixel in an inter-coded MB, we have

$$\tilde{X}(m, i, n) = \begin{cases} \hat{e}(m, i, n) + \tilde{X}(m', j, n-1) & \text{pixel received} \\ \tilde{X}(m, i, n-1) & \text{pixel lost} \end{cases} \tag{8}$$

where $\hat{e}(m, i, n)$ is the encoder-side reconstructed motion-compensated residual signal for the pixel, and $\widetilde{X}(m', j, n-1)$ is the reference pixel in frame $n-1$ pointed by the motion vector for pixel $\widetilde{X}(m, i, n)$. The two moments ($k = 1, 2$) are

$$\begin{aligned} E\{\tilde{X}(m, i, n)^k\} =& (1-p)E\{(\hat{e}(m, i, n) + \tilde{X}(m', j, n-1))^k\} \\ &+ pE\{\tilde{X}(m, i, n-1)^k\} \end{aligned} \tag{9}$$

From (6) to (9), the encoder can recursively estimate the expected distortion of each reconstructed pixel of the decoder.

3. THE PROPOSED RDE-HDA SCHEME

In this section, we generalize ROPE method to HDA-based transmission, resulting in a Recursive Distortion Estimation for HDA system (RDE-HDA). We assume that both analog signals and digital signals are affected by channel noise, and the transmission channel matches the receiver, i.e., consumers can receive all the superposed signals.

Based on the HDA framework, we can get the relationship of the digital signal, analog signal and superposition signal in encoder and decoder from Eqs. (4) and (5), respectively. Similar to ROPE method, the expected distortion of *m-th* pixel in the *i-th* block of frame *n* can be expressed as

$$\begin{aligned} ED_{HDA}(m,i,n) &= E\{(X(m,i,n)-\widetilde{X}(m,i,n))^2\} \\ &= E\{(X_a(m,i,n)-\widetilde{X}_a(m,i,n))^2\}+X_a(m,i,n)^2 \\ &+E\{(X(m,i,n)-\widetilde{X}_d(m,i,n))^2\} \\ &+2E\{X_a(m,i,n)-\widetilde{X}_a(m,i,n)\}E\{X_(m,i,n)-\widetilde{X}_d(m,i,n)\} \\ &-2E\{X_a(m,i,n)-\widetilde{X}_a(m,i,n)\}X_a(m,i,n) \\ &-2E\{X(m,i,n)-\widetilde{X}_d(m,i,n)\}X_a(m,i,n) \end{aligned} \tag{10}$$

This shows that the expected distortion in HDA is characterized by the value of $X(m,i,n)$ and the first and second moments of $X_a(m,i,n)-\widetilde{X}_a(m,i,n)$ and $X(m,i,n)-\widetilde{X}_d(m,i,n)$. $X_a(m,i,n)$ is known in the encoder, and the distortion of digital signal and analog signal can be expressed as $ED_d(m,i,n)$ and $ED_a(m,i,n)$, respectively. Similar to Eq. (6), the expected distortion of the digital signal is

$$\begin{aligned} ED_d(m,i,n) &= E\{(X(m,i,n)-\widetilde{X}_d(m,i,n))^2\} \\ &= \widetilde{X}(m,i,n)^2+E\{\widetilde{X}_d(m,i,n)^2\} \\ &-2X(m,i,n)E\{\widetilde{X}_d(m,i,n)\} \end{aligned} \tag{11}$$

Similar to Eq. (9), by assuming the transmission bit error rate is *p*, the encoder can recursively estimate the expected distortion of each reconstructed pixel of the digital signal at the decoder. The analog signal distortion $ED_a(m,i,n)$ is

$$\begin{aligned} ED_a(m,i,n) &= E\{(X_a(m,i,n)-\tilde{X}_a(m,i,n))^2\} \\ &= E\{(X_a(m,i,n)-\widehat{X}_a(m,i,n) \\ &+\widehat{X}_a(m,i,n)-\widetilde{X}_a(m,i,n))^2\} \end{aligned} \tag{12}$$

where $\widehat{X}_a(m,i,n)$ and $\widetilde{X}_a(m,i,n)$ represent analog signal after power allocation at the encoder and the corresponding noisy analog signal at the decoder, respectively. From the Softcast method we can know that each signal $\widehat{X}_a(m,i,n)$ and $\widetilde{X}_a(m,i,n)$ can be expressed as Eq. (13), according to the power allocation and transmission method in Softcast.

$$\begin{aligned} \widehat{X}_a(m,i,n) &= g_m X_a(m,i,n) \\ \widetilde{X}_a(m,i,n) &= \widehat{X}_a(m,i,n)+N(m,i,n) \end{aligned} \tag{13}$$

We divide the analog distortion into two aspects: the first is due to power allocation in Softcast, and the second is due to the transmission error. The first distortion of the *m-th* pixel in the *i-th* block of the *n-th* frame is denoted as $D_a^*(m,i,n)$, and can be expressed as

$$\begin{aligned} ED_a^*(m,i,n) &= E\{(X_a(m,i,n)-\widehat{X}_a(m,i,n))^2\} \\ &= (1-g_m)^2 E\{X_a^2(m,i,n)\} \end{aligned} \tag{14}$$

In addition, the analog signal transmission power of the *m-th* pixel in the *i-th* block of the *n-th* frame is

$$P_a(m,i,n) = g_m^2 E\{X_a^2(m,i,n)\} \tag{15}$$

From Eqs.(13) to (15), $D_a^*(m,i,n)$ can be expressed as

$$ED_a^*(m,i,n) = E\{(X_a(m,i,n) - \sqrt{P_a(m,i,n)})^2\} \tag{16}$$

Thus, the expected distortion of HDA can be expressed as

$$\begin{aligned} ED_{HDA}(m,i,n) &= E\{(X(m,i,n)-\widetilde{X}(m,i,n))^2\} \\ &= ED_d(m,i,n)+\sigma^2-2\sigma*\sqrt{ED_d(m,i,n)} \\ &-2*\sqrt{\frac{P_a}{l}}*\sqrt{ED_d(m,i,n)}+2*\sqrt{\frac{P_a}{l}}*\sigma+\frac{P_a}{l} \end{aligned} \tag{17}$$

In Eq. (17), $ED_d(m,i,n)$ can be recursively estimated through Eq. (11) in the encoder side; σ can be estimated using the standard channel estimation method for OFDM. P_a is the total power in the block of Softcast. *l* is the block size in Softcast, and equal to 64 in this paper. Thus, all the variables are known and we can accurately estimate the decoder-side distortion of HDA at the encoder side.

In addition, the theoretical BER *p* of BPSK modulation in AWGN channel is needed when estimating distortion at the encoder, and it can be described as

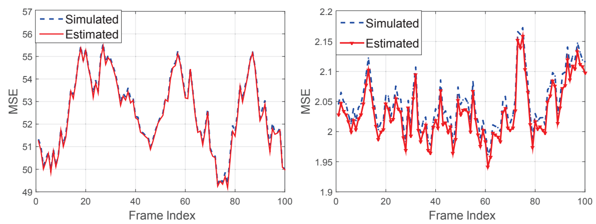
$$BER = \frac{1}{2} * erfc(\sqrt{\gamma}) \tag{18}$$

where γ is the CSNR in AWGN channel.

4. SIMULATION RESULTS

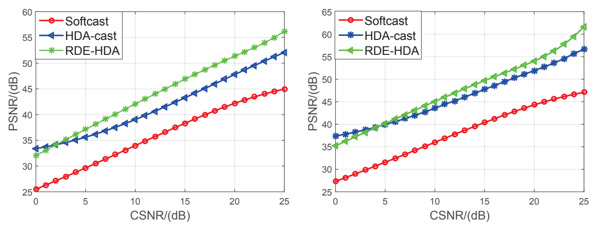
In this section, we evaluate the performance of the proposed RDE-HDA scheme by comparing the estimated distortion of our method at the encoder with the simulated distortion at the decoder. For the digital part, we employed the JM 19.0 of H.264/AVC codec to encode it. We only use single reference frame in motion estimation. The first frame is coded as I-frame, and all remaining frames are coded as P-frames. The testing sequences contain *CIF* format and *720P* format at 30 fps. The quality metric used in this paper are the Mean Squared Error (MSE), Peak Signal-To-Noise Ratio (PSNR) and the average MSE mismatch ratio (AMMR):

$$AMMR = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{ED_{est}(n) - ED_{sim}(n)}{ED_{sim}(n)}\right| \tag{19}$$



(a)*Foreman* (b)*Akiyo*

Fig. 2. Performance comparisons for the RDE-HDA method.



(a)*Foreman* (b)*Akiyo*

Fig. 3. Performance comparisons between the RDE-HDA, Softcast [2] and HDA-cast [4] with different CSNRs.

where N is the number of frames in the sequence, $ED_{est}(n)$ is the encoder-estimated MSE for frame n, and $ED_{sim}(n)$ is the receiver-side MSE for frame n obtained by simulation and averaged over multiple experiments. Only the luminance component is used in the MSE calculation.

4.1. RDE-HDA model performance

In this part, we test the performance of our proposed RDE-HDA model with different sequences. Fig. 2 shows the MSE performance of *Foreman* and *Akiyo* sequences, where the CSNR is 0 dB and 10 dB, respectively. Their AMMRs are only 0.28% and 0.32%, respectively. It can be seen that the estimated distortion at the encoder matches the actual decoder-side distortion quite well.

We also compare proposed RDE-HDA method with traditional digital scheme in **Table I**, where the CSNR=10 dB is taken as an example and only ROPE method is used in digital scheme. The results show that proposed method achieves better MSE performances meanwhile the estimated distortion is pretty precise.

4.2. AMMR for different sequences

In this part, we measure the estimation accuracy of our proposed RDE model for HDA-based transmission with more sequences and channel conditions. We choose seven sequences including *CIF* and *720P* format, and CSNRs are 3 dB and 10 dB as example. Table II presents the AMMR performance. The results show that the performance of our

Table 1. MSE Of Digital Scheme and RDE-HDA method

Sequence	Digital scheme		RDE-HDA scheme	
	Sim.	Est.	Sim.	Est.
Akiyo	19.97	19.98	3.58	3.63
Foreman	35.54	35.35	5.24	5.31
Football	39.27	41.08	5.12	5.18
Kendo	9.44	9.46	2.04	2.09

Table 2. AMMR (%) of the RDE-HDA method

Sequence	Akiyo	Balloons	Bus	Football
CSNR= 3dB	0.33	0.47	0.71	0.82
CSNR=10dB	1.60	0.73	0.86	1.21
Sequence	Foreman	News	Kendo	Average
CSNR= 3dB	0.32	0.30	1.81	0.68
CSNR=10dB	1.24	1.31	2.03	1.28

method is quite accurate. It also can be seen that the AMMR in CSNR=10 dB is higher than CSNR=3 dB. It is because that the $ED_{sim}(n)$ in CSNR=10 dB is quite smaller than C-SNR=3 dB in Eq. (18), though the average encoder-estimated MSE and receiver-side MSE is in the same level.

5. CONCLUSION

In this paper, we propose a RDE-HDA method to recursively estimate the decoder-side distortion of HDA video transmission from the encoder by allowing transmission error in the digital signal part, making HDA system more practical. Experimental results show that the accuracy of the proposed distortion estimation scheme is satisfactory.

For future work, we will consider adopting power allocation in the proposed method to reduce the distortion, and the results in this paper could provide valuable insights into decoder-side distortion estimation from a theoretical perspective, and help to guide other relevant programs which involve HDA-based transmission in the future.

6. ACKNOWLEDGEMENTS

This work has been supported in part by National Natural Science Foundation of China (No. 61672373 and No. 61501315), Program for New Century Excellent Talent in Universities (NCET-12-1037), Science and Technology Innovation Team of Shanxi Province (201705D131025), The Program of One hundred Talented People of Shanxi Province, Collaborative Innovation Center of "Internet + 3D Printing" (201708), The Key Innovation Team of the "1331 Project" of Shanxi Province (2017015).

7. REFERENCES

[1] M. Etoh, T. Yoshimura, “Advances in wireless video delivery,” *Proc. IEEE*, vol. 93, no. 1, pp. 111-122, Jan. 2005.

[2] S. Jakubczak, D.Katabi, “Softcast: one-size-fits-all wireless video.” *ACM SIGCOMM Comput. Commun. Rev.*, vol.40, no. 4, pp. 449-450, Oct. 2010.

[3] S. Jakubczak and D. Katabi, “A cross-layer design for scalable mobile video,” in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw*, Sep, 2011, pp. 289-300.

[4] L. Yu, H. Li, W. Li, “Hybrid digital-analog scheme for video transmission over wireless”, *Proc. IEEE Int. Symp. Circuits Syst.*, pp. 1163-1166, May 2013.

[5] L. Yu, H. Li, W. Li, “Wireless cooperative video coding using a hybrid digital-analog scheme”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 436-450, Mar. 2015.

[6] D. Zhang, J. Liang, I. Singh, “Fast transmission distortion estimation and adaptive error protection for H.264/AVC-based embedded video conferencing systems”, *Signal Process. Image Commun.*, vol. 28, no. 5, pp. 417-429, May 2013.

[7] Z. He, J. Cai, C. W. Chen, “Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511-523, Jun. 2002.

[8] H. Yang, K. Rose, “Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 845-856, Jul. 2007.

[9] R. Zhang, S. L. Regunathan, K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” *IEEE JSAC*, vol. 18, no. 6, pp. 966-976, July. 2000.

[10] D. Zhang, J. Liang, “View synthesis distortion estimation with a graphical model and recursive calculation of probability distribution”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 827-840, May 2015.

[11] K. Stuhlmuller, N. Farber, M. Link, B. Girod, “Analysis of video transmission over lossy channels”, *IEEE J. Select Areas Commun.*, vol. 18, pp. 1012-1032, June 2000.