

TECHNICAL RESEARCH REPORT

A Video Transmission System Based on a Human Visual Model

by Y. Jiang, J. Gu and John S. Baras

CSHCN T.R. 99-40
(ISR T.R. 99-81)



The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site <http://www.isr.umd.edu/CSHCN/>

A VIDEO TRANSMISSION SYSTEM BASED ON HUMAN VISUAL MODEL

Yimin Jiang, Junfeng Gu*, John S. Baras+*

* Hughes Network Systems, Inc, 11717 Exploration Lane
Germantown, MD 20876, E-mail: jgu, yjiang@hns.com
+ Institute for Systems Research, University of Maryland
College Park, MD 20742, E-mail: baras@isr.umd.edu

ABSTRACT

This paper presents a joint source-channel coding scheme of digital video broadcasting over satellite channel. The video compression is based on human visual model. Perceptual distortion model the just-noticeable-distortion (JND) is applied to improve the subjective quality of compressed videos. 3-D wavelet decomposition can remove spatial and temporal redundancy and provide scalability of video quality. In order to conceal the errors occurred under bad channel conditions, a novel slicing method and a joint source channel coding scenario that combines RCPC with CRC and utilizes the distortion information to allocate convolutional coding rates are presented. A new performance index based on JND is proposed and used to evaluate the overall performance at different signal-to-noise ratios (SNR). Our system uses OQPSK modulation scheme.

1. INTRODUCTION

High quality video broadcasting via satellite channel is of great interests nowadays. In this paper we focus on a satellite video transmission system that combines human visual model, 3-D wavelet subband decomposition and joint source channel coding scheme. Because the ultimate objective of video transmission systems is to maintain the subjective visual quality of images, performance metrics (other than MSE or PSNR) that take the psychovisual properties of human visual system (HVS) into account are proposed [2]. Several modern human visual models are developed, such as the just-noticeable-distortion (JND) [2] [7] and the three-component image model [5]. The JND model provides each pixel a threshold of error visibility, below which reconstruction errors are rendered imperceptible. The JND profile of a video sequence is a function of local signal properties, such as brightness, background texture, luminance changes between two frames, and frequency distribution. Scalable video compression schemes (e.g. subband coding) are widely studied [1] because they allow selective transmission of subbands to different users depending on their quality requirements and available channel bandwidths. Subband decomposition has extended to three dimensions (3-D) recently [1]. The

JND model and 3-D wavelet decomposition are applied in our video codec. The quantizer is based on the JND model and to approach the perceptual optimum.

Traditionally source and channel coders are designed independently according to Shannon's source-channel separation theorem. However in any practical communication system with finite delay and finite complexity in source and channel coders there are advantages in joint source-channel coding. [10] gives a survey of recent progress on it.

In satellite broadcast case feedback channel is not available, thus the transmitter has no information about the receivers and their channel environments. It is difficult to guarantee the average video qualities under diversified channel conditions without large channel coding overhead. We derive a new slicing method to truncate the data from each subband into small slices before arithmetic coding. Rate compatible punctured convolutional (RCPC) codes [11] are adopted in our system. The advantage of using RCPC codes is that the high rate codes are embedded into the lower rate codes of the family and the same Viterbi decoder can be used for all codes of a family. Reed-Solomon code and Ramsey interleaver plus RCPC is used to protect the data from spatial LLLL temporal L subband. Cyclic redundancy check (CRC) codes are combined with RCPC for other less significant subbands to assure acceptable video quality even under bad channel conditions.

In Section 2 the video codec is presented. Section 3 presents the error concealment scheme. In the last section the performance of the system is shown through computer simulation.

2. THE JND MODEL BASED VIDEO CODEC

Figure 1 and Figure 2 show the JND model based video encoder and decoder respectively. In the video encoder, the input video sequence is decomposed into eleven spatio-temporal frequency subbands in 3-D wavelet analysis module. The Frame Counter & Motion Detector renews the JND profiles from frame count and abrupt motion detection. The JND Model Generators estimate the spatio-temporal JND profile from analyzing local video signals and the distortion allocation algorithm that determines the JND profile for each subband. The Perceptu-

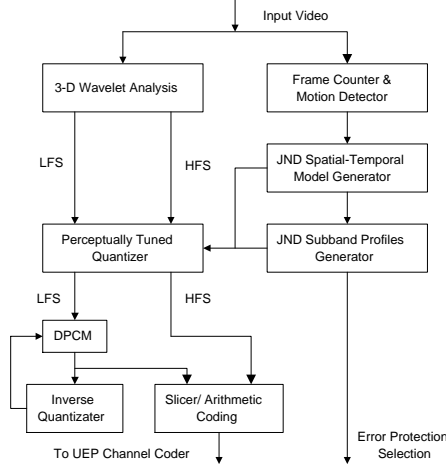


Figure 1: The JND Based Video Encoder

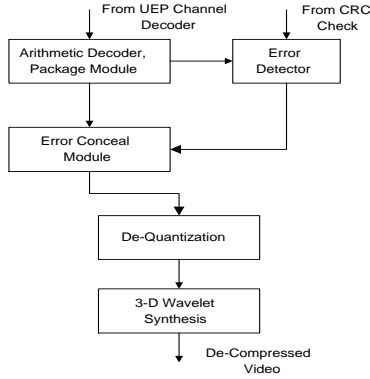


Figure 2: The JND Based Video Decoder

ally Tuned Quantizer quantizes the wavelet coefficients in each subband according to their JND profiles. The spatial LLLL temporal L subband will be encoded by DPCM. Then the data from all subbands goes through the Slicer and Arithmetic Coding part to do slicing and entropy coding. Afterward we get compressed video signal. Several modules in video codec will be presented subsequently.

2.1. 3-D Wavelet Analysis

The two-tap Haar wavelet is adopted to proceed temporal analysis, Antonini (7,9) wavelet [6] is used to proceed spatial analysis for the signal through the Haar filter. The temporal low frequency part is decomposed to two levels, and high frequency part is decomposed to one level shown as Figure 3.

2.2. The Frame Counter & Motion Detector

Because the calculation of the JND profiles is resource consuming, the Frame Counter & Motion Detector is designed to

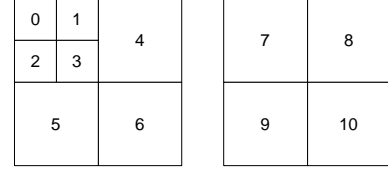


Figure 3: Subbands after 3-D Wavelet Decomposition

control the renew process of the JND. Typically the JND profiles are renewed every 10 to 20 frames, however they will be renewed immediately after an abrupt motion detected by a simple motion detector which calculates the energy of the spatial LL temporal H subband (i.e. subband 7 in Figure 3). If the energy exceeds some threshold, an abrupt motion happens with high probability.

2.3. The JND Model Generator

The JND provides each signal a threshold of visible distortion, below which reconstruction errors are rendered imperceptible. The JND profiles in spatio-temporal domain is as [3] [4], the same syntax is adopted, please refer to [3] [4] for explanation:

$$\text{JND}_{S-T}(x, y, n) \equiv f_3(\text{ild}(x, y, n)) \cdot \text{JND}_S(x, y, n) \quad (1)$$

where $0 \leq x < W$, $0 \leq y < H$, W and H are the width and height of an image, and

$$\text{JND}_S(x, y, n) \equiv \max\{f_1(\text{mg}(x, y, n)), f_2(\text{mg}(x, y, n))\} \quad (2)$$

2.4. A Novel Human Perceptual Distortion Measure

Based on the basic concept of the JND, the idea of minimally-noticeable-distortion (MND) is developed for the situation that bit-rate budget is tight and the distortion in the reconstructed image is perceptually minimal at the available bit-rate and uniformly distributed over the whole image [7]. The perceptual quality of the reconstructed image is accordingly expected to degrade evenly if bit-rate is reduced. MND is expressed as:

$$\text{MND}(x, y) \equiv \text{JND} \cdot d(x, y) \quad (3)$$

where $d(x, y)$ is the distortion index at point (x, y) . We define the energy of MND of a small area indexed by (i, j) as:

$$\sum_{(x,y) \in r_{ij}} \text{MND}^2(x, y) \equiv \sum_{(x,y) \in r_{ij}} \text{JND}^2(x, y) \cdot \delta(i, j) \quad (4)$$

where r_{ij} is a small block (typically 8 by 8 pixels), $\delta(i, j)$ is the distortion index for this block. We can define our global human perceptual distortion measure based on evaluating $\delta(i, j)$ as follows:

$$\Delta_G \equiv \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L \varepsilon(k, l) \quad (5)$$

where $\varepsilon(k, l)$ is the distortion measure of a medium block indexed by (k, l) . We decompose the whole image into K by L non-overlapped medium blocks R_{kl} ; each medium block is divided into M by N small blocks $r_{ij}(k, l)$, i.e., $R_{kl} = \bigcup_{i=1, M; j=1, N} r_{ij}(k, l)$. $\varepsilon(k, l)$ is defined as:

$$\varepsilon(k, l) \equiv \text{median}(\delta(i, j) | r_{ij}(k, l) \in R_{kl}) \quad (6)$$

The larger Δ_G is, the larger the subjective perceptual distortion. Compared with PSNR or MSE, Δ_G takes human visual model into account, therefore it can reflect the subjective visual quality better than PSNR or MSE. We will use this distortion measure to evaluate the performance of our system.

2.5. The Perceptual Tuned Quantizer

The advantage of the JND model is that it provides a quantitative measure of the error sensitivity threshold with spatial and frequency localization. Therefore the quantization table based on the JND model can be localized, which adds a globally even control on the compressed video quality. A mid-rising uniform quantizer is adopted as our basic quantizer due to its simplicity and optimal performance under certain conditions [4] [8].

First, the global object distortion index Δ_G is selected, which typically ranges from 1.0 to 10.0, where 1.0 stands for the just noticeable distortion. Second, each subband is partitioned into non-overlapped 8 by 8 blocks. For each block, the step size of the quantizer is maximized under the condition that quantization error energy is less than or equal to the MND energy (eq. (4)) of this block whose distortion index $\delta(i, j)$ equals to Δ_G . A quantization table that leads to uniform error energy over all subbands is setup for each one. It is packaged in the header of the bit stream of each subband.

2.6. Arithmetic Coding and Slicing Algorithm

Arithmetic coding [9] is adopted to achieve efficient compression, however the decoding result of one coefficient depends on the decoding result of previous one because of adaptive coding procedure. In order to prevent decoding errors from spreading, a slicing algorithm is derived to truncate the whole subband into small bit streams before arithmetic coding. The idea is to make each small bit stream carry the same amount of "distortion sensitivity". If we segment the subband S_l into Q small bit streams (or sets), each set G_i , ($i = 1, \dots, Q$) is defined as

$$G_i \equiv \left\{ (x, y) \left| \sum_{(x, y) \in G_i} \frac{1}{\text{JND}^2(x, y)} = \sum_{(x, y) \in S_l} \frac{1}{Q \text{JND}^2(x, y)} \right. \right\} \quad (7)$$

3. ERROR CONCEALMENT

A punctured code is a high rate code obtained by periodically deleting (i.e. puncturing) certain coded bits from the output stream of a low rate encoder. When compared with the optimum codes of equal rates, punctured codes are slightly less efficient but decoding complexity is greatly reduced. Clearly the puncturing rule determines the receiver structure for different code rates. A family of RCPC codes [11] are generated by adding rate-compatibility restriction to the puncturing rule. The rate-compatible restriction makes the receiver structure nearly identical for a large range of code rates. From the work in [11], we order the information bits from 11 subbands according to their source significance information (SSI) as $S_{10} \cdots S_0$. The ordered information bits are shifted into the shift registers of a $1/N$, memory M convolutional encoder. When the n_{10} information bits from subband 10 are transmitted the puncturing matrix $a(l_{10})$ (we borrow the syntax used in [11]) is used as puncturing table. As soon as the first bit from subband 9 enters the encoder the puncturing table $a(l_9)$ will be used. After another n_9 information bits, the table is switched to $a(l_8)$, etc. The procedure is easy to follow if n_k is a multiple of the puncturing period P .

In order to optimize the overall subjective video quality at reasonable coding cost, a rate allocation scheme based on JND distortion is proposed. We define the average JND distortion of subband l ($l = 0 \cdots 10$) as follows:

$$D_l = \frac{1}{H_l W_l} \sum_{(x, y) \in S_l} \text{JND}^2(x, y) \quad (8)$$

where S_l is the set of pixels of subband l , H_l and W_l are its height and weight. D_l is an indication of the robustness of S_l to quantization errors. The larger D_l is, the more robust it is to errors, the higher coding rate we choose. Table 1 shows D_l for video sequence "Mobile-Calendar". From simulation we can see that D_l divides S_l ($l = 0 \cdots 10$) into 3 groups $\{S_0\}$, $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$, $\{S_8, S_9, S_{10}\}$, which is intuitive for subband coding.

l	0	1	2	3	4	5
D_l	4.50	7.60	7.60	8.80	7.50	7.50
l	6	7	8	9	10	
D_l	8.70	7.40	27.6	27.6	42.8	

Table 1: Average Distortion D_l for Each Subband

Figure 4 shows the unequal error protection (UEP) encoder, Figure 5 shows the corresponding UEP decoder. The spatial LLLL temporal L subband (i.e. S_0) is the most significant subband, we should use a large amount of resource to protect it from errors. In order to deal with burst errors in channel Reed-Solomon (RS) code and Ramsey interleaver can be used. After RS encoder, a low rate RCPC code is selected to add

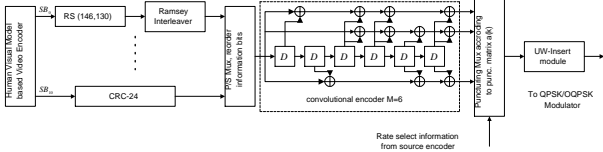


Figure 4: UEP Channel Encoder

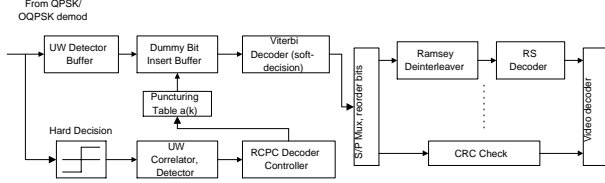


Figure 5: UEP Channel Decoder

more protection. We select a family of RCPC codes (Table II in [11], memory length $M=6$), the coding rate covers from 8/9 to 8/24. For other less significant subbands, cyclic redundancy check (CRC) codes are added before RCPC encoder to detect uncorrected errors got from channel. If errors are detected by CRC syndromes check in a block in S_l before arithmetic decoding, this block is discarded. Simulation shows that discarding an error-corrupted block is better than using wrong decoded information. Therefore the video quality degradation under bad channel conditions is acceptable because of the performance scalability of wavelet based subband coding. It gives broadcast system designers some freedom to assure the overall video quality without receivers' channel environment information at moderate coding cost. In decoder, soft-decision (or hard-decision) Viterbi decoder is adopted. An unique word (UW) detector is used to detect the start of a new subband. Rate allocation information is available on both sides.

The error probability performance of the system is discussed as following. The bit error rate upper bound of a rate $P/(P+1)$ convolutional code is given by [12]:

$$P_B < \frac{1}{P} \sum_{k=d}^{\infty} C_k P_k \quad (9)$$

where C_k is got from convolutional code generation polynomial $T(D, N)$,

$$\left. \frac{dT(D, N)}{dN} \right|_{N=1} = \sum_{k=d}^{\infty} C_k D^k$$

P_k is the probability that the wrong path with distance k to the correct path is selected. A more relax upper bound for binary symmetric channel (BSC) is given by [12]:

$$P_B < \frac{1}{P} \left. \frac{dT(D, N)}{dN} \right|_{N=1, D=\sqrt{p(1-p)}} \quad (10)$$

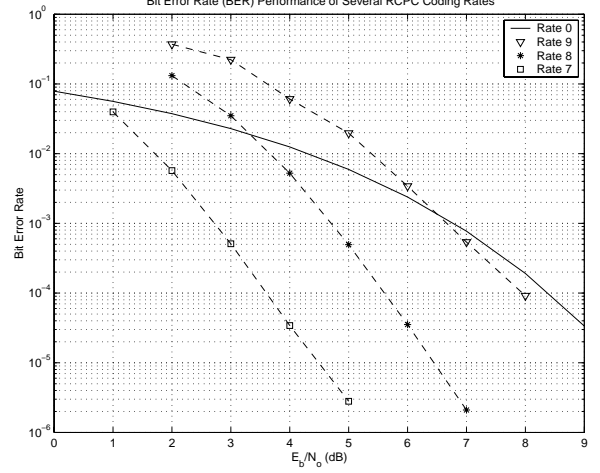


Figure 6: BER of Several RCPC Coding Rates

Index	0	1	2	3	4
Rate	1	8/24	8/22	8/20	8/18
Index	5	6	7	8	9
Rate	8/16	8/14	8/12	8/10	8/9

Table 2: Rate Index of RCPC Codes

where is p the transition probability of BSC. For OQPSK signal over AWGN channel, p is lower bounded by $Q(\sqrt{E_s/N_0})$. It is easy to achieve this bit error rate with the cutting-edge modem techniques. If CRC can detect all the channel errors, the probability that a block with length n in subband l is discarded is as follows:

$$P_D = 1 - (1 - P_B(l))^n \quad (11)$$

where $P_B(l)$ is the bit error rate of subband l with coding rate r_l if BSC model is applied.

4. SIMULATION RESULTS

Table 2 shows the index of RCPC codes used in our system, index 0 stands for no channel coding. Figure 6 shows the simulation result of the bit error rate (BER) of different coding rate at different SNR.

Video sequence "Mobile-Calendar" and "Claire" are used to test our system. Figure 7 shows Frame 1 in the recovered sequence of "Claire", which shows that the distortion measure Δ_G is better than PSNR in the sense that it reflects the subjective visual quality of image/video better. The PSNR of (a) and (b) are almost the same, but Δ_G indicates that the distortion of (a) is smaller than that of (b) as we can tell from observation (e.g. shoulder, hair and cheek). The video frames are encoded, modulated by OQPSK modulator and transmitted over AWGN channel at different SNR. The object Δ_G is set



(a)



(b)

Figure 7: Recovered Frame of "Claire" (a) $\Delta_G = 2.38$, PSNR=30.80dB (b) $\Delta_G = 3.07$, PSNR=30.15dB

to 1 which means that the compression brings just noticeable distortion in the pictures. Figure 8 and Figure 9 show distortions Δ_G and PSNR of first 10 recovered frames with different protection schemes at different SNR. The legend "Rate 4, 7, 8" means that the first UEP group (subband 0) defined previously uses the 4th RCPC code, the second UEP group (subband 1 to 7) uses the 7th RCPC code and the third UEP group (subband 8 to 10) uses the 8th RCPC code. From these figures we can see that if SNR is larger than some threshold that is different for each UEP coding scheme there will be no distortion caused by channel errors. Δ_G (Rate 4, 5, 6) is 1.22 that is larger than desired value 1 when SNR is larger than 5dB, the reason is that the JND model initially calculated from first two frames and is not renewed for the subsequent 8 frames, which brings bias in video encoding. Figure 10 shows Frame 3 of the recovered sequence of "Mobile-Calendar". Some areas corrupted by channel noise can be observed.

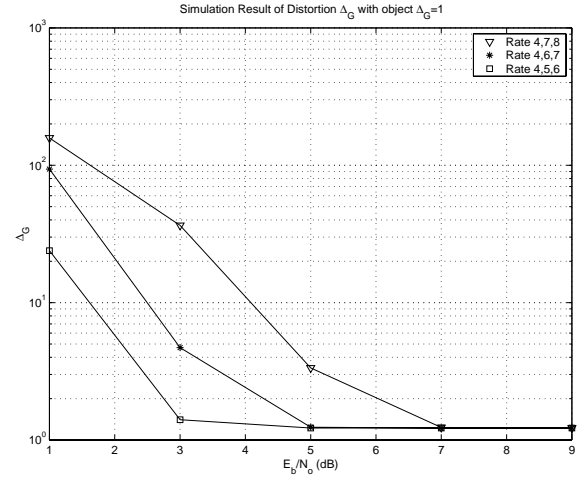


Figure 8: Δ_G with Different UEP Coding Scheme at Different SNR

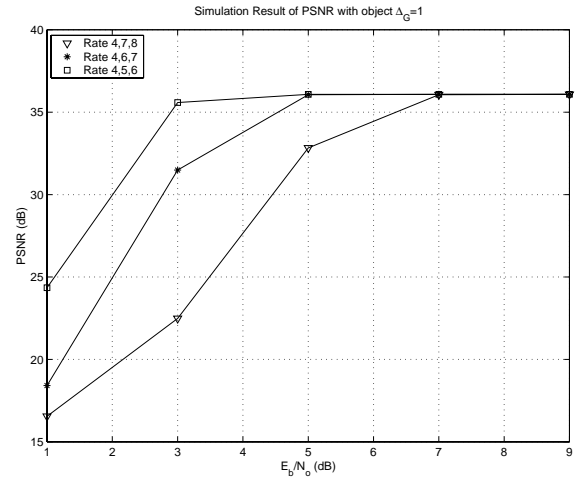


Figure 9: PSNR with Different UEP Coding Scheme at Different SNR

5. CONCLUSIONS

We have presented a satellite video transmission system based on wavelet analysis and human vision model. The joint source channel coding scheme is investigated. A new performance index based on the JND model is proposed. The quantizer and slicer are perceptually optimized. Since we focus on the subjective quality and error concealment, more powerful compression schemes (e.g. zero-tree, motion estimation and run-length coding) are not applied in our system.

6. REFERENCES

- [1] Taubman and A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. Image Processing*,



Figure 10: Recovered Frame 3 of "Mobile-Calendar" with $\Delta_G=4.235$, PSNR=30.23dB

- vol.3, pp.572-588, Sept. 1994.
- [2] N. Jayant, J. Johnston and R. Safranek, "Signal compression based on models of human perception", *Proc. IEEE*, vol.81, pp.1385-1422, Oct. 1993.
 - [3] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile", *IEEE Circuits and Systems for Video Tech.*, vol.5, pp.467-476, Dec. 1995.
 - [4] C. H. Chou and C. W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels", *IEEE Circuits and Systems for Video Tech.*, vol.6, pp.143-156, Apr. 1996.
 - [5] X. Ran and N. Farvardin, "A perceptually motivated three-component image model - Part I: Description of the model", *IEEE Trans. Image Processing*, vol.4, pp.401-415, April, 1995.
 - [6] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform", *IEEE Trans. Image Processing*, vol.1, pp.205-220, April 1992.
 - [7] N. Jayant, "Signal compression: Technology targets and research directions", *IEEE JSAC*, vol.10, pp.796-818, June 1992.
 - [8] N. Farvardin and J. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless source", *IEEE Trans. Information Theory*, vol.30, pp.485-497, May 1984.
 - [9] I. Witten, R. M. Neal and J. G. Cleary, "Arithmetic coding for data compression", *Comm. of the ACM*, Vol.30, pp.520-540, June 1987.
 - [10] N. Chaddha, S. Diggavi, "A Frame-work for Joint Source-Channel Coding of Images over Time Varying Wireless Channels ", *Proc., International Conference on Image Processing*, 1996.
 - [11] J. Hagenaur, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and Their Application ", *IEEE Trans. Comm.*, vol.36, no.4, pp.389-400 Apr. 1988.
 - [12] A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems ", *IEEE Trans. Comm.*, vol. COM-19, no. 5, pp.751- 772, Oct., 1971.